# COPULA BASED HIERARCHICAL BAYESIAN MODELS

A Dissertation

by

SOUPARNO GHOSH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Statistics

# COPULA BASED HIERARCHICAL BAYESIAN MODELS

A Dissertation

by

SOUPARNO GHOSH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

| | |
|---|---|
| Chair of Committee, | Bani K. Mallick |
| Committee Members, | Marc G. Genton |
| | Jianhua Huang |
| | Ramalingam Saravanan |
| Head of Department, | Simon J. Sheather |

August 2009

Major Subject: Statistics

# ABSTRACT

Copula Based Hierarchical Bayesian Models. (August 2009)

Souparno Ghosh, B.S., University of Calcutta;

M.S., University of Calcutta

Chair of Advisory Committee: Dr. Bani K. Mallick

The main objective of our study is to employ copula methodology to develop Bayesian hierarchical models to study the dependencies exhibited by temporal, spatial and spatio-temporal processes. We develop hierarchical models for both discrete and continuous outcomes. In doing so we expect to address the dearth of copula based Bayesian hierarchical models to study hydro-meteorological events and other physical processes yielding discrete responses.

First, we present Bayesian methods of analysis for longitudinal binary outcomes using Generalized Linear Mixed models (GLMM). We allow flexible marginal association among the repeated outcomes from different time-points. An unique property of this copula-based GLMM is that if the marginal link function is integrated over the distribution of the random effects, its form remains same as that of the conditional link function. This unique property enables us to retain the physical interpretation of the fixed effects under conditional and marginal model and yield proper posterior distribution. We illustrate the performance of the posited model using real life AIDS data and demonstrate its superiority over the traditional Gaussian random effects model. We develop a semiparametric extension of our GLMM and re-analyze the data from the AIDS study.

Next, we propose a general class of models to handle non-Gaussian spatial data. The

proposed model can deal with geostatistical data that can accommodate skewness, tail-heaviness, multimodality. We fix the distribution of the marginal processes and induce dependence via copulas. We illustrate the superior predictive performance of our approach in modeling precipitation data as compared to other kriging variants. Thereafter, we employ mixture kernels as the copula function to accommodate non-stationary data. We demonstrate the adequacy of this non-stationary model by analyzing permeability data. In both cases we perform extensive simulation studies to investigate the performances of the posited models under misspecification.

Finally, we take up the important problem of modeling multivariate extreme values with copulas. We describe, in detail, how dependences can be induced in the block maxima approach and peak over threshold approach by an extreme value copula. We prove the ability of the posited model to handle both strong and weak extremal dependence and derive the conditions for posterior propriety. We analyze the extreme precipitation events in the continental United States for the past 98 years and come up with a suite of predictive maps.

To my parents Jamuna and Nirmal Ghosh

# ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Prof. Bani K. Mallick. Without his guidance and persistent help, this dissertation would not have been possible.

I would like to thank my Ph.D. committee members, Prof. Marc Genton, Prof. Jianhua Huang, Prof. Ramalingam Saravan, for their help and support.

A special thanks goes to Prof. Michael Longnecker for bailing me out of difficulties with his invaluable advice and constant support.

A big thanks goes to my friends and peers for the enjoyable time we spent together.

# TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

# LIST OF TABLES

TABLE Page

# LIST OF FIGURES

FIGURE Page

FIGURE Page

# CHAPTER I

# INTRODUCTION

Copulas are essentially multivariate distribution functions with univariate standard uniform marginals. Thus from the basic definition, we can see, copulas provide a natural way of describing dependence among random variables. The past decade has seen a growing interest in employing copula methodology to capture the dependence structure among random variables. Consequently, copula based approaches have become a well-established tool for working with multivariate distributions.

Copulas are implicitly contained in every multivariate distribution and they do not depend on the marginal distribution of the individual components of the random vector under consideration. As a result of these two unique properties, the marginal processes and dependence structures can be modeled independently of one another. These consist of the main advantages of copula methodology.

Note that copulas require the marginals to be standard uniform distribution. But, in many situations, working with uniform marginals is neither comfortable nor justifiable. In such situations, we resort to the technique of Probability Integral Transformation (PIT) in order to transform the marginals to standard uniform distribution. However, we can show, using copulas, transforming the marginals of a random vector via PIT leaves the dependencies between the components of the random vector unchanged. In general, the dependence measures associated with copulas are nonparametric and hence can be used to capture the non-linear dependence present among the random variables. Thus, as an added benefit, the copulas provide an alternative

---

This dissertation follows the style of the *Journal of the Royal Statistical Society*.

to the Pearson correlation coefficient too. Frees and Valdez (1998), Nelsen (1999), Genest & Favre (2007) and the references therein provide a comprehensive review on this topic.

Although, copulas have found application in almost all fields to science, but their usage to study hydro-meteorological phenomena are limited. Among others, Coles & Tawn (1996a, 1996b), Dupuis (2007), Renard & Lang (2007) developed copula based methodologies to study hydrological phenomena in frequentist paradigm. But, apart from Coles & Tawn (1996b, 2005) and Sang & Gelfand (2009), no copula based models for hydro-meteorological events has been developed in the Bayesian paradigm. Neither has the copula methodology been extensively applied to study the dependence between discrete random variables. Song (2000), Li et al., (2006) adopted frequentist approach to develop Gaussian copula based model to combine discrete marginals. However, apart from the model developed by O'Brien and Dunson (2004), no Bayesian hierarchical approach employing copula technique to combine discrete marginals exists.

The main objective of our study is to employ copula methodology to develop Bayesian hierarchical models to study the dependencies exhibited by different physical processes. In particular, we concern ourselves in modeling the dependence prevalent in temporal, spatial and spatio-temporal processes. We develop hierarchical models for both discrete and continuous outcomes. In doing so we expect to address the dearth of copula based Bayesian hierarchical models to study hydro-meteorological events and other physical processes yielding discrete responses.

We now go through the content of this dissertation in greater detail. In Chapter II, we introduce the concept of copula mathematically and postulate main related results. In Chapter III, we present Bayesian methods of analysis for longitudinal binary outcomes using Generalized Linear Mixed models (GLMM). The proposed

copula model allows a very flexible marginal association among the repeated outcomes from different time-points. A unique property of this copula-based GLMM is that if the marginal link function is integrated over the distribution of the random effects, its form remains same as that of the conditional link function. The proposed models for random effects vector, viz., multivariate bridge and positive stable distributions, enable us to retain the physical interpretation of the fixed effects under conditional and marginal model. We also extend this model to a semiparametric set-up with non-parametric time-effect. We obtain several properties including the proper posterior of the fixed effects in presence of noninformative priors. Finally we illustrate our models and associated methodologies with the analysis of the longitudinal binary data from two AIDS studies.

In Chapter IV, we propose a general class of models to handle non-Gaussian spatial data. The proposed model can deal with geostatistical data that can accommodate non-Gaussianity in all its forms, viz., skewness, tail-heaviness, multimodality. We differ markedly from the earlier approaches by fixing the distribution of the marginal processes. These marginal processes are allowed to follow a non-Gaussian distribution. The spatial dependence among them is achieved via copulas (Joe, 1997; Nelsen, 1999). The idea of a copula makes our model more flexible in the sense that we allow the marginal distributions to follow any desired distribution and yet achieve dependence among them. As stated before, one of the main advantages of using copulas lies in the fact that marginal distributions may be investigated separately. The use of latent variables to transform each marginal distribution to a desired distribution is the basic tool of this modeling. We have used the multivariate elliptical distribution as the distribution of the latent variable and prove that it satisfies Kolmogorov's dimensional consistency conditions with any arbitrary marginal distribution as long as the inverse of the distribution function exists. The adoption of Bayesian approach to

perform inference about the model parameters as well as to obtain the spatial prediction at unobserved locations allows us to quantify the uncertainties associated with these estimates and predictions naturally. Furthermore, we extend this model to a mixture model framework using mixture kernels as the copula function to accommodate non-stationary data. Simulations and real data analysis show the ability of the model to identify spatial clusters. Finally, we develop a class of non-elliptical copula based models which can support a valid random field and use it to model extreme value processes.

In Chapter V, we develop copula based models for analyzing spatio-temporal dependence. In particular, we describe necessary methodologies required to model multivariate extreme values with copulas. We deal with both componentwise maxima and exceedances over thresholds. The first part of the chapter introduces the concept of extreme value copula and explains the basic methodology for studying multivariate maxima. Then we describe two traditional approaches to model maxima, viz., the block-maxima approach and the peak-over-threshold approach. We also describe how exceedances over thresholds can be approximated by an extreme value copula.

We also explore the issues of asymptotic (in)dependence in analysis of extreme values distributed over space. While asymptotically independent models are bound not to fit data that show asymptotic dependence, asymptotically dependent models can be poor approximation for asymptotically independent variables, especially for finite samples (Ledford and Tawn; 1996, 1997). In the Bayesian paradigm, the most common technique to model extreme events, distributed over space, is to assume conditional independence at the data layer (Cooley et al., 2007, Huerta and Sanso, 2007). In order to incorporate dependence information at the data level, Sang and Gelfand (2009) combined univariate extreme value distribution with a Gaussian copula and developed a valid random field with unrestricted correlation structure.

We shall show that the assumption of conditional independence at the data layer or usage of Gaussian copula to model the data layer leads to asymptotic independence. An immediate implication is that, these models cannot explain the dependence of very rare events at two specified sites, no matter how close they are. In order to circumvent this problem, we show that the proposed copula based model for extreme events is flexible enough to handle both asymptotic dependence and independence and at the same time allows unrestricted correlation structure. We illustrate our methodology by analyzing spatially distributed time series of extreme values. The model is fitted to a gridded precipitation data set collected over 99 years across the continental U.S. Predictive maps of precipitation extremes and the associated uncertainties are obtained thereafter.

# CHAPTER II

# COPULAS

In this chapter we introduce the concept of copula and describe some of its basic properties. The main textbook references on copulas are Joe (1997), Nelsen (1999) and Cherubini, Luciano & Vecchiato (2004). Throughout this dissertation, we use copulas for exploring and describing multivariate distribution function.

## II.1. **Definition and basic properties**

The joint distribution function of a random vector can be thought of as a combination of the univariate marginals of its individual components and a dependence structure describing the interactions among these marginals. Broadly speaking, copulas are functions that bind together the marginal distributions in such a manner as to form valid joint distribution function. From the formal definition below we shall see that copulas are implicitly contained in every multivariate distribution functions and can be identified with these dependence structures

*Definition 1 (copula):* A $d$ dimensional copula is the distribution function of a random vector with Uniform (0,1) marginals. The copula function $\mathbb{C} : [0,1]^d \to [0,1]$ satisfies the following properties.

(1) $\mathbb{C}(u_1, \cdots, u_d) = 0$ whenever $u_i = 0$ for at least one $i = 1, 2, \cdots, d$.

(2) $\mathbb{C}(1, \cdots, 1, u_i, 1, \cdots, 1) = u_i$ for all $i \in \{1, \cdots, d\}$, $u_i \in [0,1]$.

It is evident from the definition and the properties that any $m-$ variate marginal of a $d-$ variate copula $(2 \leq m < d)$ is itself a copula.

The Sklar's theorem (Sklar, 1959), described below, shows us how any multivariate distribution function can be described through its marginals and a copula, thereby, providing a justification for using copulas when working with multivariate distribution functions.

*Theorem 1 (Sklar):* Let $F$ be a $d-$ dimensional distribution function with margins $F_1, \cdots, F_d$. Then there exists a copula $\mathbb{C}$ such that for all $\mathbf{Y} \in [-\infty, \infty]^d$,

$$F(y_1, \cdots, y_d) = \mathbb{C}(F_1(y_1), \cdots, F_d(y_d)) \tag{2.1}$$

If $F_1, \cdots, F_d$ are all continuous, then $\mathbb{C}$ is unique; otherwise, $\mathbb{C}$ is uniquely determined on $Range(F_1) \times \cdots \times Range(F_d)$. Conversely, for a copula $\mathbb{C}$ and continuous margins $F_1, \cdots, F_d$ the function $F$ defined in (2.1) is a $d-$ dimensional distribution function with margins $F_1, \cdots, F_d$.

*Proof.* See Nelsen (1999) □

The implications of the above theorem are as follows:

(1) From the decomposition (2.1) we can study the behavior of random vectors by considering the copula and the marginals separately.

(2) Since (2.1) defines a proper distribution function for any choice of the copula $\mathbb{C}$ and marginals $F_1, \cdots, F_d$, we can construct new families of multivariate distribution with desired properties.

*Property 1.1: Uniqueness.* Let $\mathbf{Y}^{1\times d} \sim F$ with continuous marginals $F_1, \cdots, F_d$. Define the quantile function as

$$F^{-}(u) = \inf\{y | F(y) \geq u\}.$$

Since $F_i$'s are continuous and strictly increasing function, $F_i(F_i^-(y)) = y$ and by Probability Integral Transform $F_i(y) \sim$ Uniform [0,1]. According to Sklar's Theorem $F(y_1, \cdots, y_d) = \mathbb{C}(F_1(y_1), \cdots, F_d(y_d))$. Then, if $\mathbf{U} = (F_1(Y_1), \cdots, F_d(Y_d)) \sim \mathbb{C}$, then $(F_1^-(U_1), \cdots, F_d^-(U_d)) \sim F$, i.e., $\mathbb{C}(u_1, \cdots, u_d) = F(F_1^-(u_1), \cdots, F_d^-(u_d))$ is the unique copula of $F$. Another interpretation of the joint distribution function of the random vector $\mathbf{Y}$ obtained via copula $\mathbb{C}$ is the probability that each of its component is smaller than its $u_i^{th}$ quantile, independently of the marginals of its other components.

*Property 1.2: Invariance.* Let $(Y_1, \cdots, Y_d)$ be a random vector with continuous marginals and copula $\mathbb{C}$. Let $T_1 \cdots, T_d$ be strictly increasing functions $\mathbb{R} \to \mathbb{R}$. Then $(T_1(Y_1), \cdots, T_d(Y_d))$ also has copula $\mathbb{C}$.

*Proof.* See Nelsen (1999) □

The implication of this property is very important. The fact that strictly increasing transformation of the marginal distribution of a random vector does not influence the copula essentially implies that copulas capture all the information on the dependence structure without being affected by the marginal distributions.

# CHAPTER III

## BAYESIAN ANALYSIS OF LONGITUDINAL BINARY DATA WITH THE SAME MARGINAL AND CONDITIONAL LINK

In this chapter we formulate a Generalized Linear Mixed Model (GLMM) for longitudinal data, using copulas to handle the dependencies. The analysis goal is to ensure an easy interpretation of the effect of the regression variable $x(t)$ on the response $Y(t)$ at any time $t$. In the GLMM approach, the joint distribution of the response vector $\mathbf{Y} = (Y(t_1), \cdots, Y(t_m))$, measured at possibly irregularly spaced $m$ time points $t_1, \cdots, t_m$, conditional on the subject-specific mean zero vector $\mathbf{b}$ of correlated random effects, is specified by

$$\tau\{E[\mathbf{Y}|\mathbf{b}; \mathbf{X}]\} = \beta_{\mathbf{0}} + \mathbf{X}^{1\times p}\boldsymbol{\beta}^{p\times m} + \mathbf{b}^{1\times m} \ , \tag{3.1}$$

where $\tau$ is a known link function, $\mathbf{X} = (x(t_1), \cdots, x(t_m))^T$ is the regression vector, $\boldsymbol{\beta}$ is the corresponding vector of regression parameters (fixed effects) and $\beta_{\mathbf{0}} = (\beta_0(t_1), \cdots, \beta_0(t_m))^T$ is either a vector of known functions of time (with unknown parameter) or a vector of unknown intercepts. Given the subject-specific $\mathbf{b}$, the within subject responses from $m$ time points are assumed to be independent. Particularly for continuous data, we have the Linear Mixed model (LMM),

$$E[\mathbf{Y}|\mathbf{b}; X] = \beta_{\mathbf{0}} + \mathbf{X}^{1\times p}\boldsymbol{\beta}^{p\times m} + \mathbf{b}^{1\times m} \ , \tag{3.2}$$

is a special case of (3.1) with $\tau$ as the identity function.

In practice, we are often interested in the marginal regression function $E[Y(t)|X(t)]$ to understand the marginal effect of the regression coefficient $x(t)$ at time $t$ on the pop-

ulation mean response, obtained via integrating out the unobservable random effect $\mathbf{b}$. Under the LMM of (3.2), the regression parameter $\boldsymbol{\beta}$ is the same as the marginal regression parameter associated with the marginal expectation $E[\mathbf{Y}|\mathbf{X}]$. However, this is not true in general for all link functions. For example, with the Gaussian random effects for $\mathbf{b}$ in (3.1) for longitudinal binary responses (Chib and Greenberg, 1998), the marginal probability of response $E[Y(t)|x(t)] = P[Y(t) = 1|x(t)]$, after integrating over the unobservable random effects $\mathbf{b}$, in general does not follow an interpretable regression function with any familiar link function, unless we use a probit link for $\tau$. Currently popular frequentist as well as Bayesian methods of analysis of longitudinal and repeated binary outcomes using GLMMs rely heavily on either the probit link or logistic link. While the regression coefficients for the probit model have similar conditional and marginal interpretations, they do not have simple odds-ratio or relative risk type interpretations either conditionally or marginally.

We propose a class of GLMMs in which both the conditional and marginal regression parameters are easily interpretable quantities. We point out that if the distribution of $\mathbf{b}$ and the link $\tau$ are chosen properly for a GLMM in (3.1), we can ensure that

$$\tau\{E[\mathbf{Y}|\mathbf{X}]\} = \boldsymbol{\beta}_0^* + \mathbf{X}^{1\times p}\boldsymbol{\beta}^{*p\times m} \ , \tag{3.3}$$

for some link function $\tau$. The marginal regression parameters $\boldsymbol{\beta}_0^*$ and $\boldsymbol{\beta}^*$ may differ from $\beta_{\mathbf{0}}$ and $\boldsymbol{\beta}$ in (3.1), but both are readily interpretable. In particular, we can ensure that the structure of the marginal link in (3.3) to be same as the conditional link $\tau$ in (3.1), thereby preserving the physical interpretation of the regression coefficients. We present these marginally consistent models for different non-linear link functions including the logistic and complementary log-log links for binary responses and log link for count responses.

A particular choice of $\tau$ is often determined by the relevance of the link $\tau$ for useful physical interpretations of the regression effects for the practical application at hand. For example with certain studies involving binary responses, the logit is often preferred over the probit due to the log-odds interpretation of the regression effects, particularly in dose-response and exposure-response studies. As an alternative link function for binary response, a log-log link may capture the skewness of the response data better than a probit link in some applications. In essence we show that, to preserve the marginally consistent structure and simple physical interpretation of the regression coefficients in (3.1), we need not go beyond our preferred class of GLMMs. We only need to choose the distribution of $\mathbf{b}$ and the link $\tau$ judiciously. In addition to having marginal interpretations, a major advantage of our method over ordinary Bayesian GLMMs is that, under mild regularity conditions, the posterior is proper even when we use a uniform improper prior for the fixed effects.

Our methods also allow the $\mathbf{b}$ within a subject to have a very wide class of associations while the marginal density for the outcome-specific random effects will follow a certain density, for example positive stable, to assure a preferred (for example, complementary log-log link) $\tau^*$ for the marginal response in (3.3). We use a copula structure for the modeling of the vector of subject-specific $\mathbf{b}$ to guarantee a flexible longitudinal association structure as well as assuring a desired density for the marginal distribution of each component of $\mathbf{b}$. To alleviate the restriction of a completely parametric approach for modeling the time-dependent intercept term $(\beta(t_1), \cdots, \beta(t_m))$ in (3.1), we extend our methods to a class of semiparametric regression models, called partially linear models by Wang et al. (2005), Lin and Carroll (2006), where the time-dependent vector $\boldsymbol{\beta} = (\beta(t_1), \cdots, \beta(t_m))^T$ is a nonparametric function of time $(t_1, \cdots, t_m)$.

If the investigator is only interested in estimation of the marginal regression

parameters of $E[\mathbf{Y}|\mathbf{X}]$ then the generalized estimating equation (GEE) technique (Zeger and Liang, 1986; Zeger et al. 1988; Fitzmaurice et al. 1993; Fitzmaurice, 1995; Diggle et al. 2002) can be used. These GEE methods treat the within subject association, joint distribution of responses and prediction of responses over time as beyond the goals of the analysis. Due to our interest in these latter three quantities as well as the marginal regression parameters, we pursue a novel class of GLMMs.

### III.1. **Random effects model**

Our modeling goal is to define a multivariate density for $\mathbf{b}$ for the GLMM such that $\tau\{E[\mathbf{Y}|\mathbf{X}]\}$ has a linear structure while the density of $\mathbf{b}$ can accommodate a vast range of association structures among longitudinal responses measured at $m$ different time-points. Wang and Louis (2003) have presented the theory behind the class of GLMMs $\tau\{E[Y_j|b,X]\} = \beta_0 + \beta_1 X + b$ for clustered multivariate responses $\mathbf{Y} = (Y_1, \cdots, Y_m)^T$, in which the marginal response $E[Y_j|X]$ of each component $Y_j$, integrated over the common scalar random intercept $b$, can preserve the structure of the link as $\tau\{E[Y_j|X]\} = \beta_0^* + \beta_1^* X$. This is possible only using a particular random effect density $f_{B\tau}(\cdot)$ unique to the chosen link $\tau$ (called the "bridge" density of $\tau$). However, the regression parameters $\beta^*$ of the marginal function may turn out to be different from the $\beta$ of the conditional GLMM. Every link $\tau$ has its own bridge density, however, the bridge density of a link may not be unique. The general formulation for the density $f_{B\tau}$ is given as a Fourier Information transformation $f_{B\tau}(u) = (1/2\pi)\int \exp(i(k/\eta - u)v)\{\mathcal{F}(\tau(v/u))/\mathcal{F}(\tau(v))\}du$ based on the characteristic function $\mathcal{F}(\tau(v))$ of $\tau$. The multivariate model of Wang and Louis (2003) accommodates a single scalar random intercept $B$ shared by all $m$ components within a cluster, and does not allow a broader class of a vector of correlated random effects within each cluster. For longitudinal studies, the restriction to models with one scalar

random intercept shared by all responses within a subject is unappealing as associations among the repeated measures may depend on time separation. We may expect the within subject association between the responses from any two time-points to depend on the gap between time points.

We assume that the subject-specific random effects $b_{t_1}, \cdots, b_{t_m}$ at $m$ time points are different, however, they follow the same bridge density $f_{b\tau}(b|\eta_0, \eta)$ corresponding to the link $\tau$. This ensures that $\tau\{E[Y(t)|X]\} = \eta_0 + \eta(\beta_0 + X\beta)$, where $0 < \eta < 1$ is an attenuation parameter (Neuhaus et al., 1991) of the density $f_b(b_t)$. For standard links such as the logit and the log-log links, $\eta_0 = 0$ and we can also find closed form expressions of the corresponding bridge densities $f_{b\tau}$, their respective cdf $F_{b\tau}(\cdot)$ and their quantile functions $F^{-1}_{b\tau}$. In principle, we can use any $\tau$ and find its corresponding bridge density, however, we focus here only on common link functions for binary and count data to develop flexible multivariate extensions of the univariate bridge densities corresponding to these links.

To define a suitable multivariate bridge density which can accommodate a vast range of association structures within a vector $\mathbf{b}$ while ensuring the desired univariate marginal bridge density $f_{b\tau}(b)$ at any time point $t$, we use the probability integral transform (Hoel, et. al., 1971) $b_t = F^{-1}_{b\tau}(\Phi(Z_t))$, where $\Phi(\cdot)$ is the standard normal cdf, $\mathbf{Z} = [Z_{t_1}, ..., Z_{t_m}]^T$ is multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\Sigma$. The joint density of the cluster-specific $\mathbf{b}_i$ vector is given by the copula

$$f_{\mathbf{b}\tau}(b_{it_1}, \cdots, b_{it_m}) = \Phi'_{\Sigma,T}(\Phi^{-1}(F_B(b_{it_1})), \cdots, \Phi^{-1}(F_b(b_{it_m}))) \prod_{t=t_1}^{t_m} \frac{f_b(b_{it}|\eta)}{\phi\left(\Phi^{-1}(F_b(b_{it}))\right)}\ , \tag{3.4}$$

where $\Phi'_{\Sigma,m}$ represents the $m$ dimensional multivariate normal density with zero mean vector and variance-covariance matrix $\Sigma$ such that the diagonal elements equal 1 and the off-diagonal elements are $\Sigma_{ist} = Cov(Z_{is}, Z_{it})$. Since the diagonal elements equal 1,

$\Sigma_{ist} = \rho_{ist} = Corr(Z_{is}, Z_{it})$ is also the correlation between $Z_{is}$ and $Z_{it}$. Also, $\phi$ is the pdf of the standard Normal distribution and $f_b(b_{it}|\eta)$ is the univariate bridge density corresponding to the link $\tau$. For simplicity of presentation, we have suppressed in the notation the dependence of the bridge density on the corresponding link $\tau$, and we assume that all subjects are measured at time points $(t_1, \cdots, t_m)$. However, our methodology can be applied to data with a subject-specific observation schedule.

For longitudinal data, the AR(1) correlation structure

$$\rho_{ist} = \rho^{|t-s|}$$

with unknown correlation parameter $\rho$ common to every subject will be one appropriate choice for $\rho_{ist}$. In principle, any suitable correlation structure and even a subject-specific association parameter $\rho_i$ can be assumed. The linear correlation coefficient $\rho_i$ fails to capture the true dependence structure between two observable responses $(Y_i(s), Y_i(t))$. As a result, one looks for a non-parametric measure of dependence like Kendall's $\tau$ for $(Y_i(s), Y_i(t))$. One advantage of copula modeling is that, we can evaluate this relationship by calculating $\tau_{st}$, the Kendall's $\tau$ of $(Y_i(s), Y_i(t))$. Although copula functions do not have a closed form expression for $\tau_{st}$, we can calculate it via Monte Carlo simulation. Note that one advantage of the dependence measure $\tau_{st}$ is that it is independent of the regression parameter $\beta$. Thus the covariate $X_i$ of a subject does not affect the dependence measure $\tau_{st}$ within a subject. In contrast the choice of the covariate $X_i$ actually affects the linear correlation between $Y_i(s)$ and $Y_i(t)$ (Abdous et al., 2005).

The likelihood contribution $L_{M_1}(\beta, \mathbf{b}_i|Y_i)$ based on the sampling distribution of the response $Y_i = (y_i(t_1), \cdots, y_i(t_m))$ from subject $i$ is given by the joint density $\prod_{j=1}^{m_i} f(y_i(t_{ij})|\beta, \mathbf{b}_i, x_i(t_{ij}))$, where $f(y_i(t_{ij})|\beta, \mathbf{b}_i, x_i(t_{ij}))$ is based on the chosen GLMM. Assuming conditional independence of $Y_i$ for $i = 1, \ldots n$ given $(\psi, \mathbf{b}^*)$ for $\mathbf{b}^* =$

$(\mathbf{b}_1, \cdots, \mathbf{b}_n)$, we get the full likelihood as

$$L_{M_1}(\psi, \mathbf{b}|Y) = \prod_{i=1}^{n} \{L_{M_1}(\beta, \mathbf{b}_i|Y_i) f_{\mathbf{b}}(b_{i1}, \cdots, b_{it_m}|\rho, \eta)\} \ , |Y_i)\} \ , \tag{3.5}$$

where $\psi = (\boldsymbol{\beta}, \eta, \rho)$, $\quad f_{\mathbf{b}}(b_{i1}, \cdots, b_{it_m}|\rho, \eta)\}$ is the joint density of $\mathbf{b}_i$ given in (3.4) and $\eta$ is the Bridge density parameter. We further assume that the joint prior, $\pi_{M_1}(\boldsymbol{\beta}, \eta, \rho) = \pi_1(\boldsymbol{\beta})\pi_2(\eta)\pi_3(\rho)$, where the prior $\pi_1(\boldsymbol{\beta})$ for the regression parameter is typically assumed to be $MVN(\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})$. The choice of $\pi_2(\eta)$ depends on the range of $\eta$ and the available prior opinion about the variability of different clusters. The prior density $\pi_3(\rho)$ can depend on the prior opinion, when available, of the association between observables $Y_i(t)$ and $Y_i(s)$ at two different time points. A benchmark density for $\pi_3(\rho)$ is $Uniform(-1, 1)$. Under this setup, the joint posterior distribution is proportional to

$$[\prod_{i=1}^{n} \{L_{M_1}(\beta, \mathbf{b}_i|Y_i) f_{\mathbf{b}}(b_{i1}, \cdots, b_{it_m}|\eta, \rho)\}] \pi_{M_1}(\boldsymbol{\beta}, \eta, \rho) \tag{3.6}$$

Now we demonstrate how this formulation and corresponding Bayesian analysis can be achieved for different popular link functions.

### III.1.1. **Logistic link with bridge random effects**

The goal is to use a particular multivariate distribution, $f_{\mathbf{b}_i}(b_{i1}, \cdots, b_{im})$ for $\mathbf{b}_i$ such that the association structure of the multivariate density will be flexible, and the marginal regression function $E[Y_{it}|x_{it}]$ will turn out to be convenient and easy to interpret the GLM of (3.3). When $\tau$ is a logit-link, given the vector $\mathbf{b}_i$, the $Y_{it}$'s for subject/cluster $i$ are assumed to be independent Bernoulli random variables, i.e., $Y_{it}|b_{it} \sim Bern(p_{it})$, with

$$p_{it} = p_{it}(b_{it}) = \text{pr}(Y_{it} = 1|b_{it}, \mathbf{X}_{it}) = \frac{\exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})}{1 + \exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})} \ . \tag{3.7}$$

The corresponding bridge random-effects density function (Wang and Louis, 2003) is given by

$$f_b(b_{it}|\eta) = \frac{1}{2\pi}\frac{\sin(\eta\pi)}{\cosh(\eta b_{it}) + \cos(\eta\pi)}, \qquad (0 < \eta < 1\ ,\ -\infty < b_i < \infty)\ , \tag{3.8}$$

with distribution function

$$F_b(b_{it}) = 1 - \frac{1}{\pi\eta}\left[\frac{\pi}{2} - \arctan\left\{\frac{\exp(\eta b_{it}) + \cos(\eta\pi)}{\sin(\eta\pi)}\right\}\right] \tag{3.9}$$

gives us the same logit link functions for the marginal regression models, with marginal success probability

$$\text{pr}(Y_{it} = 1|\mathbf{X}_{it}) = E_b[p_{it}(b_{it})] = \frac{\exp[\eta\boldsymbol{\beta}'\mathbf{X}_{it}]}{1 + \exp[\eta\boldsymbol{\beta}'\mathbf{X}_{it}]}\ . \tag{3.10}$$

However, the conditional and the marginal regression parameters ($\boldsymbol{\beta}$ and $\eta\boldsymbol{\beta}$ respectively) are different. For simplicity, we assume the parameter $0 < \eta < 1$ of the bridge distribution to be the same at all time points.

To ensure that $b_{it}$ at any time $t$ has the marginal bridge distribution corresponding to the logit link $\tau$, we use the probability integral transform $b_{it} = F_b^{-1}(\Phi(Z_{it}))$, where $\Phi(\cdot)$ is the standard normal cdf, and

$$F_b^{-1}(u_{it}) = \frac{1}{\eta} log\left[\frac{\sin(\eta\pi u_{it})}{\sin\{\eta\pi(1-u_{it})\}}\right]$$

is the inverse cdf of that Bridge density. Thus, the $(b_{it_1}, \cdots, b_{it_m})$ within a subject are correlated as long as the $(Z_{it_1}, \cdots, Z_{it_m})$ are correlated, and the joint density of vector $\mathbf{b}_i$ is given by (3.4) with marginal cdf $F_b$ and marginal density $f_b$ given in (3.9) and (3.8), respectively.

In Figure 1, we have plotted $\tau_{st}$ versus $\rho_{st} = \rho^{|s-t|}$ for five different values of $\rho$ using this bridge density corresponding to the logit link. Note, the dependence measure $\tau_{st}$ is independent of the bridge density parameter $\eta$, and only depends on

the copula correlation $\rho$ and the distance $|s-t|$. This graph is important to understand how the dependence between $(Y_i(s), Y_i(t))$ is influenced by the correlation $\rho_{st}$ between $Z_{it}$ and $Z_{is}$. The understanding of this relationship is useful for eliciting a prior for $\rho$ when information about the association parameter $\tau_{st}$ between two observables $Y_i(s)$ and $Y_i(t)$ is available.

Since the components within the response vector $\mathbf{Y}_i$ are conditionally independent given $\mathbf{b}_i$, the likelihood contribution $L_{M_1}(\boldsymbol{\beta}, \mathbf{b}_i|Y_i)$ in (3.5) is given by $\prod_{j=1}^{m_i} p_{it_j}^{y_{ij}}(1-p_{it_j})^{1-y_{ij}}$, where $\tau(p_{it_j}) = b_{it_j} + x'_{it_j}\beta$. The choice of $\pi_2(\eta)$ depends on the available prior information about the variability of different subjects. For the MCMC computations discussed below, we have used the benchmark density $Uniform(0,1)$ for $\pi_2(\eta)$. The prior density $\pi_3(\rho)$ should depend on the prior information, when available, regarding the association between the observables $Y_i(t)$ and $Y_i(s)$ at two different time points. We have used the benchmark density $Uniform(-1,1)$ for $\pi_3(\rho)$.

*Posterior Propriety*

In a mixed effects binary regression model, improper priors on the parameters generally result in the impropriety of the posterior distribution (Natarajan and McCulloch, 1995, 1998; Natarajan and Kass, 2001; O'Brien and Dunson, 2004). Chen, Ibrahim and Shao (2004) obtained the necessary and sufficient conditions for the propriety of the posterior distribution for general classes of regression models, including the class of GLMs, under very general conditions on the model and covariates. An attractive feature of the proposed model is that, under mild conditions, which are easy to verify in practice, improper prior specification on the regression parameters leads to a proper posterior.

*Result 2.1:* Let $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n)^T$ be a sequence of binary data observed

for $n$ independent subjects arising out of the process

$$\text{pr}(\mathbf{Y}_i = y_i|\mathbf{X}_i, \boldsymbol{\beta}, \Sigma, B_i) = \prod_{t=1}^{m_i} [F(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})]^{y_{it}} [1 - F(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})]^{1-y_{it}}$$

Consider a subset of the response $\mathbf{Y}^* = (Y_{1p}, \ldots, Y_{np})$ containing only a single outcome per subject. Suppose that:

(i) The design matrix $\mathbf{X}$ is of full rank.

(ii) The likelihood of $\mathbf{Y}^*$ given $\boldsymbol{\beta}$, i.e., $L(\mathbf{Y}^*|\boldsymbol{\beta})$ has a unique maximum.

Then the joint posterior distribution arising from an improper prior of the form $\pi(\boldsymbol{\beta}, \Sigma) \propto \pi(\Sigma)$ is proper as long as $\pi(\Sigma)$ is proper.

*Proof.* See Appendix A □

In our case $\Sigma$, the correlation matrix for the random effects is characterized by $\rho$ having a compact support and hence the posterior will always be proper as long as the above two assumptions are satisfied. It is easy to verify both the assumptions in practice. The first assumption can be verified just by obtaining the determinant of the design matrix which is routinely done by all software while the second one can be checked by using any logistic regression software, which automatically checks for the existence of the MLE.

*Computation*

Following the computational scheme outlined in Zeger & Karim (1991), we implement the following MCMC steps using a Metropolis within Gibbs sampler to generate samples from the posterior.

1. For $i = 1, \ldots n$, draw $(b_{i1}, b_{i2}, \ldots, b_{im})$ from the full conditional distribution

given by

$$f(\mathbf{b}_i|\psi,\mathbf{Y}) \propto L_{M_1}(\boldsymbol{\beta},\mathbf{b}_i|Y_i) f_{\mathbf{b}}(b_{i1},\cdots,b_{iT}) \tag{3.11}$$

using the Metropolis algorithm with (3.11) as the target distribution and the proposal distribution is chosen to be Multivariate $t$ with 0 mean and degrees of freedom = 100.

2. Draw $\boldsymbol{\beta}$ using a Metropolis algorithm with the target density being the full conditional distribution given by $f_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\mathbf{b},\mathbf{Y}) \propto \prod_{i=1}^{n} L_{M_1}(\beta,\mathbf{b}_i|Y_i)\pi_1(\boldsymbol{\beta})$ with a random walk proposal.

3. Draw $\eta$ and $\rho$ similarly using a Metropolis scheme with the target distributions being $\pi_2(\eta)\prod_{i=1}^{n} f_{\mathbf{b}}(b_{i1},\cdots,b_{iT}|\eta,\rho)$ and $\pi_3(\rho)\prod_{i=1}^{n} f_{\mathbf{b}}(b_{i1},\cdots,b_{iT}|\eta,\rho)$ respectively. The proposals chosen are Uniform(0,1) and Uniform(-1,1) respectively.

Since the Bridge density has heavier tails than the Gaussian distribution, we have chosen the multivariate $t$ distribution as our proposal. From expression (3.9), we can see that if $b_{it}$ is very large or $\eta \approx 0$ then $F_b(b_{it}) \approx 1$ and consequently $\Phi^{-1}(F_b(b_{it})) \approx \infty$ so to achieve numerical stability, we perform a check that $0.00001 \leq F_b(b_{it}) \leq 0.99999$ throughout the MCMC computations.

### III.1.2. **Log-log link with positive stable random effects**

When $\tau = \log(-\log)$, we have

$$p_{it} = p_{it}(b_{it}) = \text{pr}(Y_{it}=1|b_{it},\mathbf{X}_{it}) = \exp[-\exp\{b_{it}+\mathbf{X}'_{it}\boldsymbol{\beta}\}] \ . \tag{3.12}$$

If we use $W_{it} = \exp(b_{it}) \sim St(\eta)$, which is the positive stable density with Laplace transform

$E[\exp(-uW_{it})] = \exp(-u^{\eta})$ for the heterogeneity parameter $0 < \eta < 1$, then we get the marginal probability of the response as

$$\text{pr}(Y_{it} = 1|\mathbf{X}_{it}) = \exp[-\exp\{\eta\mathbf{X}'_{it}\boldsymbol{\beta}\}] \ , \tag{3.13}$$

which has a log-log link function with attenuated regression parameter $\eta\beta$.

To ensure that $b_{it}$ has the marginal positive stable density with a flexible multivariate structure for $\mathbf{b}_i$, we again use the probability integral transform $b_{it} = F_{st}^{-1}(\Phi(Z_{it}))$, where $\Phi(\cdot)$ is the standard normal cdf, and $F_{st}^{-1}$ is the inverse of the function $F_{st}(\exp(b))$, $F_{st}(w)$ is the cdf of $St(\eta)$, and $\Phi'_{\Sigma,T}$ represents the $T$ dimensional multivariate normal density with zero mean vector and variance-covariance matrix $\Sigma$ such that $\Sigma_{st} = \rho_{st}$.

The cdf of the stable distribution does not have a closed form expression. Following Samorodnitsky & Taqqu (1994), the pdf and cdf of a random variable having a $S_{\eta}(\sigma, \beta, \mu)$ distribution can be numerically approximated by the **R** function `dstable` and `pstable` respectively. The expressions for the pdf and the cdf of a standardized stable random variable was derived by Nolan (1997) in the form of integrals. The **R** functions use these expressions and follow it up with a numerical integration to get an approximate value of the pdf and cdf of the stable random variable given all the parameters. Once we have approximated the pdf ($\tilde{f}_{W_{it}}(W_{it})$) and the cdf ($\tilde{F}_{W_{it}}(W_{it})$) of $W_{it}$, the approximate pdf and cdf of the random effect is given by $\tilde{f}_b(b^*) = \tilde{f_W}(w)\exp(b^*)$ and $\tilde{F}_b(b^*) = \tilde{F_W}(\exp(b^*))$, respectively.

Once the marginal densities of $\mathbf{b}_{it}, t = 1, 2, \ldots T$ are obtained, we can formulate the joint density of the cluster-specific $\mathbf{b}_i$ vector using the Gaussian copula. The expression of this joint density is similar to the copula in (3.4) with $F_b(b_{it})$ and $f_b(b_{it}|\eta)$ being replaced by, respectively, the cdf and pdf of the log-stable density with index parameter $\eta$, which we have approximated by $\tilde{F}_b(.)$ and $\tilde{f}_b(.)$, respectively.

Since the $Y_{it}$'s within subject $i$ are conditionally independent given $\mathbf{B}_i$, the likelihood contribution from the $i^{th}$ subject is

$$L_{M_{2i}}(\boldsymbol{\beta}, \mathbf{b}_i|\mathbf{Y}_i) = \prod_{j=1}^{m} \left[\exp(-\exp\{b_{it_j} + \mathbf{X}'_{it_j}\boldsymbol{\beta}\})\right]^{y_{ij}} \left[1 - \exp(-\exp\{b_{it_j} + \mathbf{X}'_{it_j}\boldsymbol{\beta}\})\right]^{1-y_{ij}} .$$

We use a MVN($\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}$) prior on $\boldsymbol{\beta}$, Uniform (-1,1) prior on $\rho$ and to make the marginal odds for this model comparable with the marginal odds for the logit-link model with bridge random effects and we impose a Uniform(0,1) prior on the index parameter $\eta$. Then assuming *a priori* independence of $\boldsymbol{\beta}, \rho$ and $\eta$, we get the joint prior distribution as $\pi_M(\boldsymbol{\beta}, \eta, \rho) = \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})\pi_\eta(\eta)\pi_\rho(\rho)$. The joint posterior distribution is proportional to

$$[\prod_{i=1}^{n}\{L_{M_{2i}}(\boldsymbol{\beta}, \mathbf{b}_i|\mathbf{Y}) f_{\mathbf{b}}(b_{i1}, \cdots, b_{im}|\rho, \eta)\}]\pi_M(\boldsymbol{\beta}, \eta, \rho) \tag{3.14}$$

*Computation*

1. For $i = 1, \ldots n$, draw $(b_{i1}, b_{i2}, \ldots, b_{im})$ from the full conditional distribution given by

$$f_{M_2}(\mathbf{b}_i|\psi, \mathbf{Y}) \propto L_{M_{2i}}(\boldsymbol{\beta}, \mathbf{b}_i|Y_i) f_{\mathbf{b}}(b_{i1}, \cdots, b_{iT}) \tag{3.15}$$

   using the Metropolis algorithm with (3.15) as the target distribution and the proposal distribution is chosen to be Multivariate $t$ distribution with 0 mean and degrees of freedom = 10.

2. Draw $\boldsymbol{\beta}$ using a Metropolis algorithm with the target density being the full conditional distribution given by $f_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\mathbf{b}, \mathbf{Y}) \propto \prod_{i=1}^{n} L_{M_{2i}}(\boldsymbol{\beta}, \mathbf{b}_i|Y_i)\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ with a random walk proposal.

3. Draw $\eta$ and $\rho$ similarly using a Metropolis scheme with the target distributions being $\pi_\eta(\eta) \prod_{i=1}^{n} f_{\mathbf{b}}(b_{i1}, \cdots, b_{iT}|\rho, \eta)$ and $\pi_\rho(\rho) \prod_{i=1}^{n} f_{\mathbf{b}}(b_{i1}, \cdots, b_{iT}|\rho, \eta)$, respec-

tively. The proposals chosen are Uniform(0,1) and Uniform(-1,1) respectively.

### III.1.3. **Logistic link with Gaussian random effects**

For model comparison purposes, we also present the GLMM model for binary responses with the frequently used multivariate logistic model with a multivariate Gaussian distribution for the random effects, $\mathbf{b}_i \sim \mathrm{N}_m(0, \Sigma)$, with $f^N_{\mathbf{b}_i}(b_{i1}, b_{i2}, \ldots, b_{im}) = \Phi'_{\Sigma,m}(b_{i1}, \ldots, b_{im})$ is the density of the joint multivariate normal distribution with variance-covariance matrix $\Sigma$ and mean vector zero. Under this set-up, the likelihood contribution from each subject conditional on $\mathbf{b}_i$ is given by

$$L_{M_{3i}}(\boldsymbol{\beta}, \mathbf{b}_i|\mathbf{Y}_i) = \prod_{t=1}^{m} \left[\frac{\exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})}{1+\exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})}\right]^{y_{it}} \left[1 - \frac{\exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})}{1+\exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta})}\right]^{1-y_{it}}.$$

We use a MVN($\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}$) prior on $\boldsymbol{\beta}$ and a Uniform (-1,1) prior on $\rho$. Assuming *a priori* independence of these parameters, the joint prior is given by $\pi_M(\boldsymbol{\beta}, \rho) = \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})\pi_\rho(\rho)$. So the joint posterior distribution is proportional to

$$[\prod_{i=1}^{n}\{L_{M_{3i}}(\boldsymbol{\beta}, \mathbf{b}_i|\mathbf{Y}_i) f^N_{\mathbf{b}_i}(b_{i1}, b_{i2}\cdots, b_{im}|\rho)\}]\pi_M(\boldsymbol{\beta}, \rho)$$

*Computation*

The MCMC scheme to generate samples from this joint posterior distribution is outlined below:

1. For $i = 1, \ldots n$, draw $\mathbf{b}_i$ using a Metropolis scheme with the target distribution being its full conditional distribution proportional to

$$L_{M_{3i}}(\boldsymbol{\beta}, \mathbf{b}_i|Y_i) f^N_{\mathbf{b}_i}(b_{i1}, b_{i2}, \ldots, b_{im})$$

and proposal distribution being Multivariate normal with mean $\mathbf{0}$, and Disper-

sion Matrix =100*Cov($\mathbf{Y}$).

2. Draw $\boldsymbol{\beta}$ using a Metropolis scheme with the target density being its full conditional distribution given by $g(\boldsymbol{\beta}|\mathbf{b},\mathbf{Y}) \propto \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}) \prod_{i=1}^{n} L_{M_{3i}}(\boldsymbol{\beta},\mathbf{b}_i|Y_i)$ and choosing a random walk proposal.

3. Draw $\rho$ using Metropolis scheme with target density being $g(\rho|\mathbf{b}) \propto \pi_{\rho}(\rho)$ $\prod_{i=1}^{n} f^{N}_{\mathbf{b}_i}(b_{i1}, b_{i2}, \ldots, b_{im})$ and the proposal density being Uniform (-1,1)

*Quantities of Interest*

We have denoted the logistic link model with bridge random effects as $M_1$, the log-log link model with log-stable random effects as $M_2$ and the logistic link model with Gaussian random effects as $M_3$. For all these models we report the following:

- *Deviance Information Criterion (DIC) :* Let $D(\psi)$ denote the deviance given by $D(\psi) = -2\log p(\mathbf{Y}|\psi)$ and define $\mathrm{E}(\psi) = \bar{\psi}$ and $\mathrm{E}_{\psi}(D(\psi)) = \bar{D}$. Then DIC $= 2\bar{D} - D(\bar{\psi})$. We obtain DIC for $M_1$, $M_2$ and $M_3$ and choose the model with lower value of DIC.

- *Population odds :* The within group odds is a one-to-one function of $\exp(\boldsymbol{\beta})$. We report $\exp(\boldsymbol{\beta})$ for all the considered models.

- *Marginal odds :* For $M_1$ and $M_2$ we only report $\exp(\eta\boldsymbol{\beta})$ with $\eta$ being the heterogeneity parameter for $M_1$ and the index parameter for $M_2$, while for $M_3$, we report the evaluated $E_b\left[\frac{\exp(\boldsymbol{\beta}+\mathbf{b})}{1+\exp(\boldsymbol{\beta}+\mathbf{b})}\right]$.

### III.1.4. **Example: parametric model for AIDS data**

We apply our model to a longitudinal clinical trial performed on patients infected with the human immunodeficiency virus. The purpose of this AIDS clinical trial was

to compare two therapeutic treatments, viz., zidovudine (AZT) and didanosine. The response of interest is the CD4 cell count, dichotomized at more than 200, versus, 200 or fewer cells per cubic millimeter, at weeks 0 (base-line), 2, 4 and 6. To describe the treatment effect, we form an indicator variable $\text{AZT}_i = 1$ if the $i^{th}$ subject is randomized to AZT and 0 otherwise. To control for base-line age and disease stage we define two more covariates, viz., age and AIDS. $\text{Age}_i$ is assigned the value 1 if age of the $i^{th}$ patient is 35 years or more and 0 otherwise while $\text{AIDS}_i$ is assigned the value 1 if the $i^{th}$ subject has AIDS or AIDS-related complex, and is equal to 0 if the patient is asymptomatic at base-line. The dataset contains records on 1528 patients, each of whom were supposed to be observed at 4 fixed time points corresponding to weeks 0, 2, 4 and 6. However some of the patients dropped out as the study progressed. In this article, we assume that all the missing data is non-informative.

We model the log-odds of a CD4 cell count of greater than 200 at a given time as a function of treatment, age and AIDS. Exploratory analyses showed that a time-invariant slope but time varying intercept model yields a better result. That is, we considered the following conditional logit model,

$$\begin{aligned}\text{logit}\{p_{it}\} &= \text{logit}\{\text{pr}(Y_{it} = 1|b_{it}, x_{ik})\} \\ &= \beta_0 + \beta_1\text{AZT}_i + \beta_2\text{Age}_i + \beta_3\text{AIDS}_i + \beta_4 I(t = 2) + \beta_5 I(t = 3) + \beta_6 I(t = 4)\ ,\end{aligned}$$

where $I(.)$ are indicator variables.

The results below is based on 100,000 MCMC samples with 20,000 samples constituting the burn-in period. It is noted that all the intercepts are highly significant. For the Bridge Model, the 'Age' and the 'AIDS' covariates are significant, while for the Gaussian model and the log-log link model only the 'AIDS' covariate is significant. The posterior summary of the population and marginal odds for $M_1, M_2$ and $M_3$ are

Table 1.: DICs for various models under consideration

| Parameterization | Bridge model | Log-stable model | Gaussian model |
|---|---|---|---|
| Time-invariant slope | 1617.65 | 1144.5 | 1827.34 |

shown in Figure 2, Figure 3 and Figure 4, respectively.

We have used DIC as the model selection criterion. Table 1.gives the DICs for the three models, $M_1, M_2$ and $M_3$. Note that while DIC prefers the bridge model over the Gaussian one, the log-stable random effects model with log-log link beats both the bridge model and the Gaussian model quite comfortably.

Although the log-log model performed the best among these three, it is very computationally expensive. We now extend the bridge model to a semiparametric setup .

### III.2. **Semiparametric partially linear model**

We extend the parametric model to a semiparametric setup where the effect of time has been assumed to be unknown and estimated using spline basis function. Suppose for individual $i = 1, \ldots, n$, the random follow-up times are denoted by $\{T_{i1} < \ldots < T_{im_i}\}$. Let $\{Y_i(t)\}$ with $Y_{ij} = Y_i(T_{ij})$ denote the longitudinal outcome process with the $j^{th}$ observed follow-up time for the $i^{th}$ individual denoted by $t_{ij}$ and the corresponding observed longitudinal response denoted by $y_{ij}$. For a given individual, define $\mathbf{N} = [N_1, \ldots, N_T]^T$ to be the $T \times 1$ random vector with $N_t$ taking the value 1 if $Y(t)$ is observed at time $t$ and taking the value 0 otherwise. Clearly $\sum_{j=1}^{T} N_{ij} = m_i$. Thus, the process $\{\mathbf{N}_i(t)\}$ equals the number of follow-up visits of subject $i$ by time $t > 0$. In this paper, we assume that the stochastic model of $\{\mathbf{N}_i(t)\}$ can be ignored in order to draw inferences on the model parameters (Ryu et al., 2007). Let the values of

$Y(t)$ at the observed follow-up times be denoted by $\mathbf{Y}_o$. Then, under this 'ignorable' follow-up model, we can write the likelihood as

$$L(\Theta|\mathbf{Y}_o, \mathbf{X}) = \prod_{i=1}^{n} \prod_{t=1}^{m_i} f_{Y_i(t_{it})}(y_{it}|Y_{i1} = y_{i1}, \ldots, Y_{i(t-1)} = y_{i(t-1)}) \ . \tag{3.16}$$

The longitudinal model of $\{Y_i(t)\}$ is given by

$$\text{Logit}\left[\text{P}(Y_{it} = 1|\mathbf{X}_{it}, b_{it})\right] = g(t) + \mathbf{X}'_{it}\boldsymbol{\beta} + b_{it} \ , \tag{3.17}$$

where $\mathbf{b}_i = (b_{i1}, \ldots, b_{im_i})^T$ has a multivariate bridge distribution whose density is derived in (3.4). The correlation structure is assumed to be AR(1). We use a Bayesian natural cubic regression spline to model $g(t)$. With a certain number $p$ of knots, $t_1 < \ldots < t_p$, with $t_j \in [T_{\min}, T_{\max}]$, we consider the space $\mathcal{S}_q\ (t_1, \ldots, t_p)$ of splines of order $q$. Here we take $q = 4$ to get the natural cubic splines. In this space, $\mathcal{S}_4(t_1, \ldots, t_p)$ we can represent $g$ by

$$g(t) = \sum_{j=1}^{p} C_j S_j(t) = \mathbf{S}(t)\mathbf{C} \tag{3.18}$$

with known cubic spline basis functions $\mathbf{S}(t) = [S_1(t), \ldots, S_p(t)]$ and unknown basis coefficients $\mathbf{C} = (C_1, \ldots, C_p)^T$.

So the contribution of the $i^{th}$ subject in the Likelihood conditional on the parameters is given by

$$L(\mathbf{Y}_i|\boldsymbol{\beta}, \mathbf{C}, \mathbf{b}_i) = \prod_{t=1}^{m_i} \left[\frac{\exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta} + g(t))}{1 + \exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta} + g(t))}\right]^{y_{it}} \left[1 - \frac{\exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta} + g(t))}{1 + \exp(b_{it} + \mathbf{X}'_{it}\boldsymbol{\beta} + g(t))}\right]^{1-y_{it}} \ .$$

Then combining (3.16) and (3.17) we get the full likelihood to be

$$L(\mathbf{Y}_o|\Theta, \mathbf{b}) = \prod_{i=1}^{n} L(\mathbf{Y}_i|\boldsymbol{\beta}, \mathbf{C}, \mathbf{b}_i) \ , \tag{3.19}$$

where $\Theta = (\boldsymbol{\beta}, \eta, \rho, \mathbf{C})$. We use a time-varying intercept but a time-invariant slope

regression model for the parametric part.

The joint prior distribution is assumed to be $\pi_s(\Theta) = \pi_s(\boldsymbol{\beta}, \eta, \rho, \mathbf{C}) = \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})\pi_\eta(\eta)$ $\pi_\rho(\rho)\pi_{\mathbf{C}}(\mathbf{C})$ where $\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ is MVN($\mathbf{0}$, $\Sigma_{\boldsymbol{\beta}}$), $\pi_\eta(\eta)$ is Uniform (0,1), $\pi_\rho(\rho)$ is Uniform (-1,1) and $\pi_{\mathbf{C}}(\mathbf{C})$ is $N_p(0, I_p)$.

Under this set-up the joint posterior is proportional to :

$$\left[\prod_{i=1}^{n} L(\mathbf{Y}_i|\boldsymbol{\beta}, \mathbf{C}, \mathbf{b}_i) f_{\mathbf{b}}(b_{i1}, \cdots, b_{im_i}|\rho, \eta)\right] \pi_s(\boldsymbol{\beta}, \eta, \rho, \mathbf{C}) \ .$$

*Computation*

With the formulation in (3.18), the obvious question is the number of knots and location of the knots. A computationally rigorous way to answer this question is to employ a Reversible Jump MCMC to choose the optimum number of knots and then optimally choose their location. To circumvent this computational complexity, we use a fixed-knots approach and select the final model using DIC. The MCMC scheme to draw samples from the joint posterior distribution is described below:

1. For $i = 1, \ldots n$, draw $(b_{i1}, \ldots, b_{im_i})$ from the full conditional distribution given by

$$f(\mathbf{b}_i|\Theta, Y_i) \propto L(\mathbf{Y}_i|\boldsymbol{\beta}, \mathbf{C}, \mathbf{b}_i) f_{\mathbf{b}}(b_{i1}, \cdots, b_{im_i}) \tag{3.20}$$

   using a Metropolis algorithm with (3.20) as the target distribution and the proposal distribution is chosen to be Multivariate $t$ with 0 mean and degrees of freedom = 100.

2. Draw $\mathbf{C}$ from the full conditional distribution given by

$$f_{\mathbf{C}}(\mathbf{C}|\mathbf{b}, \mathbf{Y}) \propto \pi_{\mathbf{C}}(\mathbf{C}) L(\mathbf{Y}_o|\Theta, \mathbf{b})$$

   with a random walk proposal.

3. Draw $\boldsymbol{\beta}$ using Metropolis algorithm with the target density being the full conditional distribution given by $f_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\mathbf{b},\mathbf{Y}) \propto L(\mathbf{Y}_o|\Theta,\mathbf{b})\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ with a random walk proposal.

4. Draw $\eta$ and $\rho$ similarly using a Metropolis scheme with the target distributions being $\pi_\eta(\eta)\prod_{i=1}^n f_{\mathbf{b}}(b_{i1},\cdots,b_{im_i})$ and $\pi_\rho(\rho)\prod_{i=1}^n f_{\mathbf{b}}(b_{i1},\cdots,b_{im_i})$, respectively. The proposals chosen are Uniform(0,1) and Uniform(-1,1), respectively.

### III.2.1. **Example: semiparametric model for AIDS data**

We use the same clinical trial as in Section 3 to explore the Semiparametric model. The response of interest is still the CD4 cell count. However, in this section, we use all of the collected outcome data, which includes 15 time points (basline and weekly through 14 weeks from randomization). Further, along with the two covariates and the treatment AZT, we add the covariate 'performance status' (PERF90) to the regression model. The performance status is an attempt to quantify AIDS patients' general well-being, with PERF90=1 if the performance status is good, and PERF90=0 if the performance status is poor. We assume an 'ignorable' follow-up model. Exploratory analyses yielded better results when we modelled the parametric part with time-varying intercepts but a time-invariant slope regression model. However the effect of time is taken into account in the nonparametric part, where fixed points in time are used to form the spline basis functions.

We perform 100,000 iterations and use the first 20,000 iterations as the burn-in period. We find all the intercepts and the four covariates, viz., PERF90, AIDS, Age and AZT, to be significant. For the population-level odds and marginal odds, we report only $\exp(\beta)$ and $\exp(\eta\beta)$ respectively. The posterior summaries of these representative values of the odds are shown in Figure 5.

Figure 6 shows a comparison between the performance of the bridge random effects semiparametric model developed in (3.17) as compared to the bridge random effects ordinary parametric model described in (3.7) for two representative subjects, viz., Subject 5 and Subject 29. The chosen two subjects are representative in the sense that, while observations corresponding to every time-point for Subject 5 are zero and thus show no curvature at all; Subject 29 shows a well balanced occurrence of zeros and ones distributed over the considered time-period, thereby showing good curvature. We note that the semiparametric model successfully captures the curvature of the probabilities of success but the ordinary parametric model assigns almost the same probability of success at all time-points. This explains why DIC prefers the semiparametric model, with its DIC value equal to 4.03, to the ordinary parametric one whose DIC value is 11.40.

Figure 7 shows the plot of Kendall's $\tau$ of $(Y_a, Y_{a+t})$ at different lags for different values of $\rho$.

### III.3. **Concluding remarks**

We have developed parametric and semiparametric joint models for longitudinal multivariate binary data. This model can easily be extended to a situation where there is an additional continuous outcome of interest that is also measured repeatedly over time. For example, in longitudinal studies of cardiac patients, repeated measurements of the binary outcome 'shortness of breath' (yes/no) and the continuous outcome systolic blood pressure are often collected. In such a situation, we can extend the results to a joint longitudinal model of a binary and a continuous outcome at each time point. For a joint analysis of both outcomes, the binary outcome can be modeled as in Section III.1 and the continuous outcome can be modeled using a random effects model. The correlation between the longitudinal binary and continuous outcomes can

be induced by specifying correlations between the continuous random effect and the bridge random effect using a joint multivariate Gaussian distribution.

# CHAPTER IV

# BAYESIAN MODELING OF NON-GAUSSIAN GEOSTATISTICAL DATA VIA COPULAS

Spatial dependence is as common and as interesting as temporal dependence. But temporal processes have been explored to a far greater extent than the spatial processes. Consequently, our understanding of the former is greater than that of the latter. Relying on that understanding, we have developed a copula based Generalized Linear Mixed Model for multivariate binary data observed over time, in the previous chapter (Chapter III). It will be interesting to explore how we can adapt that methodology to study spatial dependence. In this chapter, we develop a very general model for analyzing spatial data, using copulas, and show that many traditionally used spatial models are special cases of this copula-based geostatistical model.

Modeling spatial data with Gaussian processes is the common thread of all geostatistical analyses. However, non-Gaussian characteristics, such as nonnegative continuous variables with a skewed distribution, often with a heavy right or left tail or with multiple modes, appear in many data sets from scientific fields. We need ways to model those kinds of data sets. A common way to model this type of data is to assume that the random field of interest is the result of an unknown nonlinear transformation of a Gaussian random field. Trans-Gaussian kriging is the kriging variant used for prediction in transformed Gaussian random fields, where the normalizing transformation is assumed known. This approach has some potential weaknesses (De Oliveira et al., 1997; Azzalini et al., 1999) such as (1) the transformations are usually on each component separately, and achievement of joint normality is only hoped for;

(2) the transformed variables are more difficult to interpret, especially when each variable is transformed by using a different function; (3) even though the normalizing transformation can be estimated by maximum likelihood, it may be unwise to select a particular transformation.

Alternatively, more general (flexible) parametric classes of multivariate distributions can be used to represent features of the data set as adequately as possible and to reduce unrealistic assumptions. The pioneering work was started by Zellner (1976) who proposed the regression model with multivariate Student t error terms. Many interesting classes of distributions are reviewed by Johnson and Kotz (1972). Moore flexible classes of sampling models (Palacios and Steel, 2006; Kim and Mallick, 2003) have been developed based on a scaled mixture of Gaussian processes. A scale mixture of Gaussian processes is successful in modeling heavy tailed symmetric spatial data but may fail to capture the skewness present in the data. Recently, Kim and Mallick (2004) introduced a skew-Gaussian process mimicking the characterization of a Gaussian process. The characterization of this skew-Gaussian process is improper as it does not advocate a valid stochastic process. Hence, predicted values of the process at unobserved locations are not self coherent.

In this article we propose a novel way of dealing with geostatistical data that can accommodate non-Gaussianity in all its forms, viz., skewness, tail-heaviness, multimodality. We differ markedly from the earlier approaches by fixing the distribution of the marginal processes. These marginal processes are allowed to follow a non-Gaussian distribution. The spatial dependence among them is achieved via copulas (Joe, 1997; Nelsen, 1999). Recently, a Bayesian approach based on Gaussian copulas had been developed by Pitt et al. (2006). The idea of a copula makes our model more flexible in the sense that we allow the marginal distributions to follow any desired distribution and yet achieve dependence among them. The use of latent variables to

transform each marginal distribution to a desired distribution is the basic tool of this modeling. We have used the multivariate elliptical distribution as the distribution of the latent variable and prove that it satisfies Kolmogorov's dimensional consistency conditions with any arbitrary marginal distribution as long as the inverse of the distribution function exists. We adopt a fully Bayesian approach to perform inference about the model parameters as well as to obtain the spatial prediction at unobserved locations. Furthermore, we extend this model to a mixture model framework using mixture kernels as the copula function to accommodate non-stationary data. Simulations and real data analysis show the ability of the model to identify spatial clusters. Finally, we develop a class of non-elliptical copula based models which can support a valid random field and use it to model extreme value processes.

## IV.1. **Methodology**

### IV.1.1. **Overview of elliptical distribution**

A $n \times 1$ random vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ is said to have an elliptical distribution with mean $\boldsymbol{\mu}$ and covariance $\Sigma$ if its density function is given by

$$|\Sigma|^{-\frac{1}{2}} g\left((\mathbf{Z}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{Z}-\boldsymbol{\mu})\right) \tag{4.1}$$

for some function $g$ satisfying the following properties:

1. $g$ must be non-negative
2. $g(||t||), \quad t \in \mathbb{R}^k$ is a characteristic function, where $||t||$ denotes the norm of a vector $t$.
3. $$g(u) = \int_{[0,\infty)} \Omega_k(r^2 u) dF(r) \quad u \geq 0$$

for some distribution function $F$ on $[0, \infty)$, where $\Omega(||t^2||), \quad t \in \mathbb{R}^k$, is the characteristic function of a $k$-dimensional random vector $U^k$ which is uniformly distributed on the unit sphere in $R^k$. Detailed description of these properties can be found in Cambanis et al. (1981) and the references therein.

We use the notation $\mathbf{Z} \sim EC_n(\boldsymbol{\mu}, \Sigma, g)$ and the pdf of $\mathbf{Z}$ as obtained in (4.1) is denoted by $g_{\boldsymbol{\mu},\Sigma}(Z_1, Z_2, \ldots, Z_n)$.

### IV.1.2. **Properties of elliptical distribution**

We note down a few properties of elliptical distributions that are required to build a spatial model.

1. All distributions in the class $EC_n(\boldsymbol{\mu}, \Sigma, g)$ have the same mean $\boldsymbol{\mu}$ and same correlation matrix.

2. All marginal distributions of $\mathbf{Z}$ are identical with density function denoted by $q_g(.)$ and distribution function denoted by $Q_g(.)$

3. All marginal density functions of dimension $j < n$ are elliptical and have the same functional form.

4. For any given $(m \times n)$ matrix $D$ of rank $m, (m \leq n)$, the random vector $D\mathbf{Z} \sim EC_m(D\boldsymbol{\mu}, D\Sigma D^T, g)$.

5. If $\mathbf{Z} \sim EC_n(\boldsymbol{\mu}, \Sigma, g)$ is partitioned as $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)^T$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^T$ and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

with $\mathbf{Z}_1$ and $\boldsymbol{\mu}_1$ are $k \times 1$ vector and $\Sigma_{11}$ is a $k \times k$ matrix, then the conditional distribution of $\mathbf{Z}_1|\mathbf{Z}_2$ is a $k-$variate elliptical distribution with mean and

covariance given by

$$\begin{aligned} E(\mathbf{Z}_1|\mathbf{Z}_2) &= \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{Z}_2 - \boldsymbol{\mu}_2) \\ Cov(\mathbf{Z}_1|\mathbf{Z}_2) &= h(\mathbf{Z}_2)\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \end{aligned}$$

for some function $h$ depending on the exact pdf of $\mathbf{Z}$.

### IV.1.3. **Formulation of spatial model**

Let $Y(\mathbf{S})$ be a random process defined for locations $\mathbf{S}$ in a spatial region $D \in R^n$. Usual assumption is $Y(\mathbf{S})$ is a second-order stationary error process with mean 0 and has a valid correlation function of distance between the location points, parameterized by a vector $\theta$. We assume that we have observed a single realization from this random process at $n$ different locations $\mathbf{S}_1, \cdots, \mathbf{S}_n$, and we denote the vector of observations by $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$ where we use the notation $Y_i = Y(\mathbf{S}_i)$. Usually a joint distribution is assumed for $\mathbf{Y}$. In the literature, by far the most commonly made stochastic assumption is that $\mathbf{Y}(\mathbf{S})$ is a Gaussian process. In contrast, we specify the marginal distribution of $Y_i$ and construct the joint distribution by using copula.

Let $q_g(.)$ and $Q_g(.)$ be respectively the marginal pdf and cdf of an $n$ dimensional elliptical distribution $EC_n(0, \Sigma, g)$. Let each component $Y_i$ of $\mathbf{Y}$ have absolutely continuous distribution with density function denoted by $f_i(y_i)$ and distribution function denoted by $F_i(y_i)$. Assuming that $Q_g^{-1}(.)$ exists, we get the copula density for $\mathbf{Y}$ as:

$$\begin{aligned} f_{\mathbf{Y}}(y_1, \ldots, y_n|\Sigma, \boldsymbol{\eta}) &= |\Sigma|^{-\frac{1}{2}} g_{0,\Sigma}\left(Q_g^{-1}(F_1(y_1;\eta_1)), \cdots, Q_g^{-1}(F_n(y_n,\eta_n))\right) \\ &\times \prod_{i=1}^{n} \frac{f_i(y_i;\eta_i)}{q_g(Q_g^{-1}(F_i(y_i;\eta_i)))} \end{aligned} \tag{4.2}$$

where $\eta_i$ is the vector of parameters for the CDF $F_i(y_i)$ and $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)$.

In process of modeling the joint distribution, we have to model the copula func-

tion $\mathbb{C}(.)$ and the individual marginal distribution $F_i$, $i = 1, \cdots, n$. Since, the copula density involves the usage of an elliptical density function, we need to specify the covariance function $\Sigma$ for developing the spatial process. We assume the covariance function only depends on the distance between two locations and not on the direction (isotropy). For simplicity, we assume a generalized exponential covariance function $\Sigma_{\boldsymbol{\theta}}(l) = \exp(-\nu l^{\theta_2}) = \theta_1^{l^{\theta_2}}$ where $l$ is the Euclidean distance between locations, $\nu > 0, \quad \theta_1 = \exp(-\nu)$ and $\boldsymbol{\theta} = [\theta_1, \theta_2]$. That is, for two distinct locations $S_i$ and $S_j$ separated by the distance $l_{ij}$, we assume $\text{Cov}(Q_g^{-1}(F_i(y_i, \eta_i)), Q_g^{-1}(F_j(y_j, \eta_j)))= \Sigma_\theta(l_{ij})$. The assumption of generalized exponential covariance structure also allows us to compare our results with the existing methods. However, extension to use more general class of covariance functions is conceptually straightforward. In next section we shall develop some flexible spatial processes by proper selection of copula structure and marginal processes and discuss some spatial assumptions imposed on the marginal parameters $\boldsymbol{\eta}$.

**Flexible spatial processes**

A spatial location $\mathbf{S}_i$ is characterized by its spatial coordinates $(X_{i1}, X_{i2})$. Define $\mathbf{X}_i = (1, X_{i1}, X_{i2})$ and $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]^T$. We assume a second order stationary process with isotropic covariance structure. Furthermore, for simplicity, we assume that the mean surface $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T$ is linear. Since, we want to incorporate covariates in the marginal processes and model the skewness explicitly, we assume that the marginal processes belong to the skew-elliptical family (Genton, 2004) with location parameter $\mu_i$, scale parameter $\rho_i$, and shape parameter $\alpha_i$. We relate these marginal parameters with additional covariates in the process layer of the model.

*Gaussian process and Gaussian anamorphosis models*

If we use a Gaussian copula with Gaussian marginals for $\mathbf{Y}$, the model reduces to the conventional zero-mean Gaussian process. The Gaussian copula with the copula density is given by

$$f_{\mathbf{Y}}(y_1,\ldots,y_n|\Sigma,\boldsymbol{\eta}) = \phi_{0,\Sigma}\left(\Phi^{-1}(F_1(y_1;\eta_1)),\cdots,\Phi^{-1}(F_n(y_n,\eta_n))\right)\times\prod_{i=1}^{n}\frac{f_i(y_i;\eta_i)}{\phi(\Phi^{-1}(F_i(y_i;\eta_i)))} \tag{4.3}$$

where $\phi(.)$ and $\Phi(.)$ are the density function and distribution function of a standard normal distribution.

Furthermore, using a Gaussian copula with any other continuous marginal distribution for $\mathbf{Y}$ yields the Bayesian model based approach with Gaussian anamorphosis structure (Chilès & Delfiner, 1999).

*Heavy tailed spatial processes*

Similarly, we can generate heavy tailed spatial processes (Palacios and Steel, 2006) using t copula or Logistic copula. The copula density for the $t$ copula is given by

$$f_{\mathbf{Y}}(y_1,\ldots,y_n|\Sigma,\boldsymbol{\eta}) = t_{\Sigma,k}\left(T_k^{-1}(F_1(y_1;\eta_1)),\cdots,T_k^{-1}(F_n(y_n,\eta_n))\right)\times\prod_{i=1}^{n}\frac{f_i(y_i;\eta_i)}{t_k(T_k^{-1}(F_i(y_i;\eta_i)))} \tag{4.4}$$

where $t_{\Sigma,k}$ denote the p.d.f of an $n$-variate $t$-distribution with covariance matrix $\Sigma$ and $k$ degrees of freedom and $T_k$ and $t_k$ denote the distribution function and density function of an univariate $t$-distribution with d.f $k$ and variance 1.

*Logistic copula*

We followed O'Brien & Dunson (2004) proposal of a new parameterization of multivariate logistic distribution via multivariate $t$-distribution which is given by

$$\begin{aligned} L_\nu(z_1, z_2, \ldots, z_n) &= t_{\Sigma,\nu}\left(T_\nu^{-1}\left(1/1+e^{-z_1}, \eta_1\right), \cdots, T_\nu^{-1}\left(1/1+e^{-z_n}, \eta_n\right)\right) \\ &\quad \times \prod_{i=1}^{n} \frac{L_1(z_i;\eta_i)}{t_\nu(T_\nu^{-1}(F_i(y_i;\eta_i)))} \end{aligned} \tag{4.5}$$

where $t_{\Sigma,\nu}$, $T_\nu$ and $t_\nu$ are defined analogously as in $t$ copula and and $L_1(z_i, \eta_i)$ is the density function of the univariate logistic distribution with parameter $\eta$. The authors further showed that this multivariate logistic distribution can be almost exactly approximated by setting $\nu = 7.3$. In this paper, we approximate the logistic copula density function by the expression (4.4) with $k = 7.3$

*Skewed spatial processes*

We develop skew-Gaussian processes by using Gaussian copula as in equation (4.3) and fixing the marginal distribution of $Y_i$ to be skew normal (Azzalini and Capitanio, 1999; Azzalini and Dalla Valle, 1996; Genton, 2004) with the density function given by:

$$f_{sn}(y_i; \mu_i, \rho_i, \alpha_i) = \frac{2}{\rho_i}\phi\left(\frac{y_i-\mu_i}{\rho_i}\right)\Phi\left(\alpha_i\frac{y_i-\mu_i}{\rho_i}\right),\ \infty \le y_i \le \infty \tag{4.6}$$

where $\phi(.)$ and $\Phi(.)$ are described earlier. Recently, Kim and Mallick (2004) introduced skew Gaussian process mimicking the characterization of Gaussian process. The characterization of their skew-Gaussian process is improper as they do not advocate a valid stochastic process hence predicted values of the process at unobserved locations are not self coherent.

Furthermore, we develop heavy tailed skewed spatial processes by fixing the marginal distribution of $Y_i$ to be skew-$t$ and combining the marginals with the ellip-

tical copulas. The density function of skew-$t$ is given by (Genton, 2004)

$$f_{st}(y_i;\mu_i,\rho_i,\alpha_i) = 2f_{k_1,k_2}(y_i;\mu_i,\rho_i)F_{k_1^*,k_2^*}(\alpha_i \times (y_i-\mu_i)) \tag{4.7}$$

where $f_{k_1,k_2}(.;\mu_i,\rho_i)$ is the pdf of a univariate generalized $t$ distribution with location, scale and shape parameters are given by $\mu_i$, $\rho_i, k_1$ and $k_2$, respectively, and $F_{k_1^*,k_2^*}(.)$ is the cdf of a univariate standard generalized $t$ distribution, with $k_1^* = k_1 + 1$ and $k_2^* = k_2 + \frac{(y_i-\mu_i)^2}{\rho_i}$. For simplicity we fix, $k_1 = k_2 = 1$. Similarly, development of skew Laplace copula is straight forward. Thus, the marginal parameters $\eta_i$s are characterized by the site-specific location parameter $\mu_i$, scale parameters $\rho_i$ and the shape parameters $\alpha_i$.

### IV.1.4. **Hierarchical model**

In this section we shall discuss the development of the hierarchical Bayesian models and implementation issues using MCMC computation.

*Data model*

Given the entire parameter vector $\xi_1 = (\boldsymbol{\mu}, \boldsymbol{\rho}, \boldsymbol{\alpha}, \boldsymbol{\theta})$, where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, using (4.2), we obtain the joint distribution of $\mathbf{Y}$ as

$$\begin{aligned} L(\mathbf{Y}|\xi_1) &\propto g_{0,\Sigma}\left(Q_g^{-1}(F_{Y_1}(y_1;\mu_1,\rho_1,\alpha_1)),\ldots,Q_g^{-1}(F_{Y_n}(y_n;\mu_n,\rho_n,\alpha_n))\right) \\ &\times \prod_{i=1}^{n} \frac{f_{Y_i}(y_i;\mu_i,\rho_i,\alpha_i)}{q_g\left(Q_g^{-1}(F_i(y_i;\mu_i,\rho_i,\alpha_i))\right)}. \end{aligned} \tag{4.8}$$

This constitutes the data layer of our model. We need to verify that the model proposed in (4.8) indeed supports a valid stochastic process satisfying Kolmogorov's conditions. Hence the following results :

*Result 3.1:* $L(\mathbf{Y}|\boldsymbol{\eta})$ is absolutely continuous.

*Result 3.2:* $L(\mathbf{Y}|\boldsymbol{\eta})$ supports a valid stochastic process.

*Result 3.3:* If the correlation structure for $Q_g^{-1}(F(y))$ is isotropic, so is the dependence structure of $\mathbf{Y}$.

*Proof.* See Appendix A □

*Process model*

In this layer of hierarchy, we relate marginal parameters of the data layer with available covariates in a hierarchical fashion. Furthermore, we model the mean function as a Gaussian process with its mean depending on the covariates and its covariance depending upon the distance between the sites. Accordingly, we have $\boldsymbol{\mu}|\boldsymbol{\rho} \sim N(X^T\beta_\mu, diag(\rho_1,\ldots,\rho_n)\Sigma_{\theta_\mu})$ where $\Sigma_{\theta_\mu} = \theta_{1\mu}^{l^{\theta_{2\mu}}}$. We model scale parameters as $\log(\boldsymbol{\rho}) = (\log\rho_1,\ldots,\log\rho_n) \sim N(-0.5, \Sigma_{\theta_\rho})$ where $\Sigma_{\theta_\rho} = \theta_{1\rho}^{l^{\theta_{2\rho}}}$. We have a little information about the sensitivity of the shape parameters to the regional covariates. However, it is fair to assume that they will depend on the spatial lag between two locations. Hence, we assume $\boldsymbol{\alpha} \sim N(0, \Sigma_{\theta_\alpha})$ where $\Sigma_{\theta_\alpha} = \theta_{1\alpha}^{l^{\theta_{2\alpha}}}$. Finally, we assume a Uniform (0,1) prior for $\theta_1$ and Uniform (0,2] prior for $\theta_2$. Denoting $\xi_2 = (\beta_\mu, \theta_{1\mu}, \theta_{2\mu}, \theta_{1\rho}, \theta_{2\rho}, \theta_{1\alpha}, \theta_{2\alpha})$, we get the process model as

$$\pi(\xi_1|\xi_2) \propto f(\boldsymbol{\mu}|\beta_\mu, \boldsymbol{\rho}, \theta_{1\mu}, \theta_{2\mu}) \times f(\boldsymbol{\rho}|\theta_{1\rho}, \theta_{2\rho}) \times f(\boldsymbol{\alpha}|\theta_{1\alpha}, \theta_{2\alpha}).$$

*Priors*

In this final hierarchical structure we assign priors on $\xi_2$. We assume $\beta_\mu \sim N(0, \Sigma_\mu I)$ for some large $\Sigma_\mu$. $\theta_{1\mu}, \theta_{1\rho}, \theta_{1\alpha}$ are assumed to be independently distributed as Uniform (0,1) and $\theta_{2\mu}, \theta_{2\rho}, \theta_{2\alpha}$ are distributed independently as Uniform (0,2]. Thus, the prior model is given by

$$\pi(\xi_2) \propto f(\beta_\mu).$$

Using the hierarchical structure described above, we get the joint posterior distribution proportional to $L(\mathbf{Y}|\xi_1) \times \pi(\xi_1|\xi_2) \times \pi(\xi_2)$.

### IV.1.5. **Prediction**

Let $\mathbf{Y}_0^{k\times 1} = \mathbf{Y}(S_0) = (\mathbf{Y}(S_{01}, \ldots, \mathbf{Y}(S_{0k}))$ be the realizations at unobserved locations $(S_{01}, \ldots, S_{0k})$. In order to predict $\mathbf{Y}_0$, we need to calculate the posterior predictive density given by :

$$f(\mathbf{Y}_0|\mathbf{Y}) = \int f(\mathbf{Y}_0|\xi, \mathbf{Y})\pi(\xi|\mathbf{Y})d\xi \tag{4.9}$$

where $\xi = (\xi_1, \xi_2]$ is the entire set of parameters . Further, we assume that $\mathbf{Y}_0$ comes from the same random field as do the observations $\mathbf{Y}$. Let us denote

$$\Sigma^{*(n+k\times n+k)} = \begin{pmatrix} \Sigma^{n\times n} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}^{k\times k} \end{pmatrix}$$

to be the dispersion matrix corresponding to the augmented data vector $(\mathbf{Y}, \mathbf{Y}_0)^T$. Further, we set $u_i^* \stackrel{\Delta}{=} F_{Y_i}(y_i; \mu_i, \rho_i, \alpha_i)$. Then from (4.8), we get the joint distribution of $(\mathbf{Y}, \mathbf{Y}_0|\xi)$ to be

$$f(\mathbf{Y}_0, \mathbf{Y}|\xi) \propto g_{0,\Sigma^*}(Q_g^{-1}(u_1^*), \ldots, Q_g^{-1}(u_{n+k}^*))\frac{\prod_{i=1}^{n+k} f_{Y_i}(y_i; \mu_i, \rho_i, \alpha_i)}{\prod_{i=1}^{n+k} q_g(Q_g^{-1}(u_i^*))}. \tag{4.10}$$

Using (4.10) and (4.8), we get the conditional distribution of $\mathbf{Y}_0$ given $\mathbf{Y}$ and $\xi$ as

$$\begin{aligned} f(\mathbf{Y}_0|\mathbf{Y}, \xi) &\propto \frac{g_{0,\Sigma^*}(Q_g^{-1}(u_1^*), \ldots, Q_g^{-1}(u_{n+k}^*))}{g_{0,\Sigma}(Q_g^{-1}(u_1^*), \ldots, Q_g^{-1}(u_n^*))} \prod_{i+1}^{n+k} \frac{f_{Y_i}(y_i; \mu_i, \rho_i, \alpha_i)}{q_g(Q_g^{-1}(u_i^*))} \\ &= R \prod_{i+1}^{n+k} \frac{f_{Y_i}(y_i; \mu_i, \rho_i, \alpha_i)}{q_g(Q_g^{-1}(u_i^*))} \end{aligned} \tag{4.11}$$

where

$$R = \frac{g_{0,\Sigma^*}(Q_g^{-1}(u_1^*), \ldots, Q_g^{-1}(u_{n+k}^*))}{g_{0,\Sigma}(Q_g^{-1}(u_1^*), \ldots, Q_g^{-1}(u_n^*))}. \tag{4.12}$$

From Property 5 in Section IV.1.2, we can see that $R$ is the density function of a $k$ variate elliptical distribution. For example, in case of Gaussian copula, $R$ is the density of a $k$ variate Normal distribution with mean $E_k = -\Sigma_{21}\Sigma^{-1}\mathbf{V}_1$ and variance $\Gamma_k = \Sigma_{22} - \Sigma_{21}\Sigma^{-1}\Sigma_{12}$ where $\mathbf{V}_1 = (\Phi^{-1}(u_1^*), \ldots, \Phi^{-1}(u_n^*))$.

Denoting the conditional mean and conditional variance of the $k$ variate elliptical distribution by $E_k$ and $\Gamma_k$ respectively, we can write :

$$R = g_{E_k,\Gamma_k}(Q_g^{-1}(u_{n+1}^*), \ldots, Q_g^{-1}(u_{n+k}^*)). \tag{4.13}$$

Combining (4.13) and (4.11), we get the conditional distribution of $\mathbf{Y}_0$ given $\mathbf{Y}$and $\boldsymbol{\eta}$ as

$$f(\mathbf{Y}_0|\mathbf{Y}, \xi) \propto g_{E_k,\Gamma_k}\left(Q_g^{-1}(u_{n+1}^*), \ldots, Q_g^{-1}(u_{n+k}^*)\right) \prod_{i+1}^{n+k} \frac{f_{Y_i}(y_i; \mu_i, \rho_i, \alpha_i)}{q_g(Q_g^{-1}(u_i^*))}. \tag{4.14}$$

*Computation*

As the joint posterior distribution cannot be analyzed analytically, we have to rely on Markov chain Monte Carlo (MCMC) methods to simulate the parameters from this posterior distribution. Furthermore, the full conditional distributions are not of explicit form. Hence, we use Metropolis-Hastings algorithm to simulate all the parameters. The detail steps of MCMC computations are in the supplementary website.

Finally, we develop an algorithm to approximate the posterior predictive density $f(\mathbf{Y}_0|\mathbf{Y})$ as follows.

1. Discretize the effective range of $\mathbf{Y}_0$ to get the set $S_0$.

2. Generate $\xi_{11}, \ldots, \xi_{1s} \sim$ iid $\pi(\xi_1|\xi_2)$. Note that conditional on $\xi_2$, the elements of $\xi_1$ are independent. So this amounts to drawing the data layer parameters from their respective priors.

3. For all $\mathbf{Y}_0 \in S_0$ approximate $f(\mathbf{Y}_0|\mathbf{Y})$ by

$$\widehat{f}_s(\mathbf{Y}_0|\mathbf{Y}) = \sum_{i=1}^{s} f(\mathbf{Y}_0|\mathbf{Y}, \xi_{1i})\omega(\xi_{1i}) \tag{4.15}$$

where

$$\omega(\xi_i) = \frac{f(\mathbf{Y}|\xi_{1i})}{\sum_{j=1}^{s} f(\mathbf{Y}|\xi_{1j})}$$

and $f(\mathbf{Y}_0|\mathbf{Y}, \xi_{1i})$ is given in (4.14) and $f(\mathbf{Y}|\xi_{1i})$ is given in (4.8).

Geweke (1989) had shown that under regularity conditions $\widehat{f}_s(\mathbf{Y}_0|\mathbf{Y})$ is a consistent estimator of $f(\mathbf{Y}_0|\mathbf{Y})$ and $\widehat{f}_s(\mathbf{Y}_0|\mathbf{Y}) \rightarrow^{a.s} (\mathbf{Y}_0|\mathbf{Y})$ as $s \rightarrow \infty$. Once we obtain the posterior predictive distribution of $\mathbf{Y}$, the median of $f(\mathbf{Y}_0|\mathbf{Y})$ yields a robust estimate of the response at locations $S_0$ which is denoted as $\widehat{\mathbf{Y}_0}$.

### IV.1.6. **Model adequacy**

Since the main purpose of the proposed model is prediction, we use a cross validation approach based on single-point-deleted predictive distribution to assess the adequacy of our model. Let $y_{i,obs.}$ be the observed value at $i$th location and $y_{-i,obs.} = (y_{1,obs.}, \ldots, y_{i-1,obs.}, y_{i+1,obs.}, \ldots, y_{n,obs.})$ be the data set with $i$ th observation deleted, $i = 1, \ldots, 24$. To assess the predictive accuracy of our models, we need to ascertain how well the models predict each $Y_i$ based on $y_{-i,obs.}$. If $\hat{Y}_i$ denotes the median of the predictive distribution $(Y_i|y_{-i,obs.})$, then a natural measure to assess the predictive accuracy is obtained by the prediction residual given by $r_i = (y_{i,obs.} - \hat{Y}_i)$. Once we get the prediction residuals, then we can compare the predictive accuracy of the various competing models using the mean square prediction error (MSPE) criterion given by $\frac{1}{n}\sum_{i=1}^{n} r_i^2$ and the empirical coverage probability of the nominal 95% prediction intervals.

### IV.1.7. **Simulation: stationary random field**

We perform extensive simulations to evaluate the performance of proposed models under complete misspecification. We generate data from the random field using Gaussian, $T_2$ and Logistic copulas. For each copula, we select two different type of marginal distributions, viz., skew-normal with $\rho = 12$ and $\alpha = 2.5$ and skew-$t_2$ with $\rho = 8.3$ and $\alpha = 5$. Hence, we have six copula-marginal combinations of true models. The correlation matrix, $\Sigma_\theta$, has the generalized exponential structure defined earlier with $\theta_1 = 0.5, \theta_2 = 1.7$.

Then we fit these six copula-marginal combination models to the generated data sets to compare their performance under model misspecification. For this and all other subsequent analyses, the degrees of freedom for the $t$ copula and skew-$t$ marginals are obtained via exploratory analyses. We check the model adequacy and evaluate the predictive performance using the methods described in Section IV.1.6. The predictive performance in terms of MSPE and coverage probability of 95% predictive interval (shown in parenthesis) are shown in Table 2..

It is evident that in general the predictive performances of the models with correctly chosen marginals are better than those with misspecified marginals. This indicates that the choice of marginals is more crucial from the prediction perspective. Moreover, for most of the cases, the empirical coverage of the 90% prediction intervals, though lower than the nominal level, are fairly stable and not overly optimistic. Finally, the $T_2$-copula-skew $t$-marginals model has consistently performed better than the other proposed models and their performance is the most robust under model misspecifications.

Table 2.: Comparison among various copula models

| **Fitted Model** | **True Model** | | | | | |
|---|---|---|---|---|---|---|
| | Gaussian Skew-normal | Gaussian Skew-$t$ | $T_2$ Skew-normal | $T_2$ Skew-$t$ | Logistic Skew-normal | Logistic Skew-$t$ |
| Gaussian Skew-normal | 0.9301 (91%) | 3.1881 (78%) | 0.8281 (89%) | 3.1584 (77%) | 1.3911 (74%) | 3.4427 (72%) |
| Gaussian Skew-$t$ | 0.9318 (94%) | 1.5976 (94%) | 0.8717 (94%) | 1.3776 (94%) | 1.4630 (88%) | 1.9586 (89%) |
| $T_2$ Skew-normal | 1.009 (95%) | 2.1875 (87%) | 0.7657 (94%) | 2.7749 (80%) | 1.1430 (84%) | 3.2732 (78%) |
| $T_2$ Skew-$t$ | 1.0205 (96%) | 1.7389 (94%) | 0.7611 (96%) | 1.2941 (96%) | 1.2005 (95%) | 1.8944 (95%) |
| Logistic Skew-normal | 1.2725 (89%) | 3.2021 (79%) | 0.7852 (85%) | 2.8463 (74%) | 0.9476 (93%) | 3.0033 (75%) |
| Logistic Skew-$t$ | 1.2894 (95%) | 1.6483 (90%) | 0.8252 (89%) | 1.5396 (93%) | 0.9547 (95%) | 1.3532 (93%) |

### IV.1.8. **Example: modeling spatial rainfall pattern**

The dataset comprises of rainfall amounts (in mm) accumulated over a 7-day period, from 76th day to the 82nd day of 1991 near Darwin, Australia during the rainy season. The rainfall was measured using rain gauges at 24 sites located in a region, called a D-scale region, of about 10 km $\times$ 10 km. The D-scale network was designed approximately on a regular 10 km $\times$ 10 km grid with gauge spacing approximately 2 km. The purpose of this D-scale network is to provide high quality validation data over a small domain. However, site availability limited the establishment of gauges, and the average gauge density is one gauge per 5.3 km$^2$. The main purpose of this study is to devise a model which can be used for accurate spatial interpolation, thus alleviating the problem of relative sparsity of rain gauges in the D-scale region. If the proposed models demonstrate significantly accurate predictive capability, they can be used for spatial interpolation to obtain high quality validation data in spite

of low gauge density in the D-scale region. As a preliminary adjustment, we center the observations around the observed mean and then scale it down by the observed standard deviation. The D-scale region is located on the coastal pain of the Adelaide river where the terrain is very flat with no orographic or climatological variations, hence we assume the covariates to be the latitude and longitude of a site under consideration. We consider three types of copulas, namely, Gaussian , $t$ and logistic and two types of marginals, namely, skew-Gaussian and skew-$t$ yielding six competing models each with different copula-marginal combination. The degrees of freedom for the $t$ copula is chosen based on exploratory analyses. The values for $\Sigma_\mu$ has been varied from 5 to 100 to assess prior sensitivity.

We compare the predictive accuracy of our model with that of the Bayesian Trans-Gaussian kriging (BTG), Trans-Gaussian Kriging (TGK), log-normal kriging (LNK) and ordinary kriging (OK) methods. Table 3. provides the MSPE and coverage probability of the 95% prediction interval for various copula-marginal combinations and other geostatistical models.

It can be clearly seen that the copula models perform better than the other existing models with the $t$ copula-skew $t$ marginal model showing the best performance. Figure 8(a) shows the plot of the posterior median of $\boldsymbol{\alpha}$ along with its 95% credible interval observed at different sites in the D-scale region using the best fitted model. Note that for most of the sites, the credible interval for $\alpha$ does not contain 0 indicating departure from Gaussianity.

Finally, as a rough measure to assess the goodness of fit of our model, we plot the kernel smoothed densities of the actual observations along with the kernel smoothed predictive density in Figure 8(b). It seems that the predictive distribution fits the observed data adequately.

In the following section we discuss how our proposed models can be modified to

Table 3.: Comparison of copula models with kriging variants

| Model/ Copula | Marginals | MSPE | 95% PI coverage |
|---|---|---|---|
| Gaussian | Skew-normal | 41.24 | 79.16% |
| | Skew-$t$ | 38.75 | 87.5% |
| $T_{10}$ | Skew-normal | 36.15 | 87.5 % |
| | Skew-$t$ | **27.90** | **91.6** % |
| Logistic | Skew-normal | 41.86 | 83.3% |
| | Skew-$t$ | 46.38 | 83.3% |
| BTG | - | 51.73 | 91.6% |
| TGK | - | 77.99 | 75% |
| OK | - | 58.45 | 79.1 % |
| LNK | - | 68.56 | 75% |

handle binary data and illustrate this modification with a real life example.

## IV.2. **Modeling spatial binary data**

The proposed copula method can be easily extended to model discrete data. We illustrate this extension by modeling a realization of binary data obtained over a spatial random field. The key idea of this binary-copula model is to write it in the form of a generalized linear mixed effects model. While the fixed effects capture the mean surface, the random effects are introduced to capture the underlying spatial process.

Assuming a logit-link and given the vector of random effects $\boldsymbol{b} = (b_1, \ldots, b_n)$, the response $Y_i$ for location $i$ is assumed to be independent Bernoulli random variable, i.e., $Y_i|b_i \sim Bern(p_i)$, with

$$p_i = \text{pr}(Y_i = 1|b_i, \boldsymbol{\beta}) = \frac{\exp(b_i + \mathbf{X}_i'\boldsymbol{\beta})}{1 + exp(b_i + \mathbf{X}_i'\boldsymbol{\beta})} \tag{4.16}$$

where $b_i$ follows an univariate skew elliptical distribution with mean 0, scale $\rho_i$ and shape $\alpha_i$. These marginal densities are combined using an elliptical copula to yield a valid spatial process. Then the likelihood is given by

$$L(Y|\boldsymbol{\beta}, \Sigma, \boldsymbol{\rho}, \boldsymbol{\alpha}) \propto \int \prod_{i=1}^{n} \text{pr}(Y_i = 1|\boldsymbol{\beta}, b_i) f(\boldsymbol{b}|0, \Sigma) d\boldsymbol{b}$$

where $f(\boldsymbol{b}|0, \Sigma)$ is joint density of $\boldsymbol{b}$ described in (4.2) with $\Sigma = \theta_1^{l^{\theta_2}}$. In the second level of hierarchy, we assume $\boldsymbol{\beta} \sim N_k(0, \Sigma_\beta)$, $\log(\boldsymbol{\rho}) \sim N_n(-0.5, \Sigma_{\theta_\rho}(l))$, and $\boldsymbol{\alpha} \sim N_n(0, \Sigma_{\theta_\alpha}(l))$, where $\Sigma_{\theta_\rho}(l) = \theta_{1\rho}^{l^{\theta_{2\rho}}}$ and $\Sigma_{\theta_\alpha}(l) = \theta_{1\alpha}^{l^{\theta_{2\alpha}}}$. We assume $\theta_1$ is distributed as Uniform (0,1), while $\theta_2$ is distributed as Uniform (0,2]. In the third layer of hierarchy, we impose priors on $\Sigma_\beta, \theta_{1\rho}, \theta_{1\alpha}, \theta_{1\rho}, \theta_{1\alpha}$. We assume the parameters to be independent *a priori* and assign an Inverse-Wishart$(I, 10)$ prior on $\Sigma_\beta$, $\theta_{1\rho}, \theta_{1\alpha}$ are assumed to distributed independently as Uniform (0,1), while $\theta_{2\rho}, \theta_{2\alpha}$ are distributed independently as Uniform (0,2]. Then the joint posterior distribution is proportional to

$$f(\boldsymbol{\beta}|0, \Sigma_\beta) f(\boldsymbol{\rho}| - 0.5, \Sigma_{\theta_\rho}) f(\boldsymbol{\alpha}|0, \Sigma_{\theta_\alpha}) f(\Sigma_\beta) \prod_{i=1}^{n} \text{pr}(Y_i = 1|b_i, \boldsymbol{\beta}) f(\boldsymbol{b}|0, \Sigma).$$

We draw samples from the posterior distribution using Metropolis within Gibbs sampling scheme.

### IV.2.1. **Example: outbreak of equine encephalomyelitis in Texas**

An outbreak of equine West Nile Virus encephalomyelitis cases were reported to Texas disease authorities during 2002. This had a significant effect on the equine population of Texas. Since, the horses form a significant proportion of livestock in the state of Texas, the economic impact of this outbreak was pronounced on the state rural economy. A standard equine neurological disease report was completed recording outcomes (recovered, died/ euthanasized) of 1299 cases diagnosed with West Nile

Virus encephalomyelitis, their spatial location, the clinical symptoms associated with the disease and the vaccination status. Using this data, we intend to explore the spatial variation present in the odds of a case of death (versus surviving). In doing so, we try to identify significant disease clusters and predict the odds of death at unobserved locations.

Let $Y_i = Y(S_i)$ denote the outcome of a particular case at location $S_i$ which takes the value 0 if the horse survived and takes the value 1 otherwise. Then conditional on the underlying stationary spatial process $b(S)$, the model for $Y_i$ can be written as a classical generalized linear mixed model given by

$$\text{logit}[E(Y_i|b_i)] = \mathbf{X}_i'\boldsymbol{\beta} + b_i.$$

The covariates available for this study are the clinical symptoms including ataxia, falling-down, recumbency, lip-droop. The vaccination status of the horse as well as the spatial coordinate at which the case was reported are other covariates.

*Results and Model Selection*

Since the main focus of this model is to predict the odd of death of a horse at an unobserved location, we perform a leave-one-out cross validation as outlined in Section IV.1.6. As a measure of performance, we use the percentage of correctly classified cases with respect to survival status based on the single-point-deleted predictive distributions. The predicted odd of death at various locations are shown in Figure 9. A disease cluster can easily be identified in the north-eastern part of Texas.

Table 4. shows the classification accuracy of various models obtained assuming different spatial processes for $\boldsymbol{b}$. Note that all copula variants have better classification accuracy as compared to the traditional Gaussian random effects model.

Table 4.: Classification performance of various copula models

| Model/ Copula | Marginals | Percentage of observations correctly classified |
|---|---|---|
| Gaussian | Skew-normal | 0.8815 |
| | Skew-$t$ | 0.9037 |
| $T_{10}$ | Skew-normal | 0. 8855 |
| | Skew-$t$ | 0.8926 |
| Logistic | Skew-normal | 0.8845 |
| | Skew-$t$ | 0.8906 |
| Gaussian | - | 0.80 |

## IV.3. **Modeling non-stationary random field**

Here we propose an extension to the copula model discussed so far so that it can accommodate covariance-non-stationary spatial processes, and thereby can identify homogeneous data clusters. This model is rich enough to accommodate multi modal, skewed and heavy-tailed data.

For this purpose, we propose a mixture of elliptical kernels as our copula function.We define a $M^{th}$ order mixture model as a model where the number of components in the mixture is $M$. In the following sections, we describe the full hierarchical structure of the $M^{th}$ order mixture model.

### IV.3.1. **Hierarchical model**

*Data model*

Let the structural covariance parameter vector $\boldsymbol{\theta}$ be partitioned into $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{21}, \ldots, \theta_{M1})$ and $\boldsymbol{\theta}_2 = (\theta_{12}, \theta_{22}, \ldots, \theta_{M2})$ and the mixing parameters be denoted by $\Pi = (\pi_1, \ldots, \pi_M)$. Then, conditional on the parameters $\xi_1 = (\boldsymbol{\mu}, \boldsymbol{\rho}, \boldsymbol{\alpha}, \theta_1, \theta_2, \Pi)$,

the data layer model is given by:

$$L_M(\mathbf{Y}|\xi_1) \propto \sum_{j=1}^{M} \pi_j g_{0,\Sigma_j}\left(Q_g^{-1}(F_1(y_1;\mu_1,\rho_1,\alpha_1)),\ldots,Q_g^{-1}(F_n(y_n;\mu_n,\rho_n,\alpha_n))\right) \prod_{i=1}^{n} \frac{f_i(y_i;\mu_i,\rho_i,\alpha_i)}{q_g\left(Q_g^{-1}(F_i(y_i;\mu_i,\rho_i,\alpha_i))\right)}. \tag{4.17}$$

where $\Sigma_j = (\Sigma_{\theta_j}(l)) = \theta_{j1}^{l^{\theta_{j2}}}$.

*Result 3.4:* Properties of single component model (Result 3.1-3.3) hold true for the sampling model described in (4.17).

*Proof.* See Appendix A . □

Note that the data model described in (4.17) is symmetric under all permutations of $\pi$ and $\Sigma_j$, thus giving rise to the so called 'label switching' problem. However, the main focus of the present study is spatial prediction, therefore we are more concerned with the posterior predictive distribution. In the later section, we have derived the predictive density under this sampling model, and that does not depend on how the mixture components are labeled. The fact that the predictive density is unaffected by the label switching problem was also noted by Stephens (2000). However, for inferential purpose, we do impose identifiability constraints to circumvent the label switching problem. Simulation studies and exploratory analyses that we have undertaken prompted us to impose the following identifiability constraints:

$$\theta_{11} > \theta_{21} > \ldots \theta_{M1} \text{ and } \theta_{12} > \theta_{22} > \ldots \theta_{M2}.$$

*Process model*

Once again, in this layer we relate the parameters of the data layer with all available covariates associated with the spatial process. We assume that $\boldsymbol{\mu} \sim N(h(\beta_\mu, \mathbf{X}), \Sigma_{\theta_\mu})$

for some function $h(.)$. We model log-transformed scale parameters as $\log \boldsymbol{\rho} \sim N(-0.5, \Sigma_{\theta_\rho})$. The shape parameters are allowed to follow a $N(0, \Sigma_{\theta_\alpha})$ distribution where the structures of $\Sigma_{\theta_\mu}, \Sigma_{\theta_\rho}, \Sigma_{\theta_\alpha}$ are same as that described in Section IV.1.4. The mixing parameters $\Pi$ are assumed to follow Dirichlet $(\gamma_1, \gamma_2, \ldots, \gamma_n)$. Under the identifiability conditions described earlier, we propose a structured model for $\boldsymbol{\theta}$. We assume a single step Markovian prior structure for the joint distribution of both $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, that is,

$$\pi(\boldsymbol{\theta}_1) = \pi(\theta_{11})\pi(\theta_{21}|\theta_{11}) \ldots \pi(\theta_{M1}|\theta_{(M-1)1})$$
$$\text{with } \theta_{11} \sim Uniform(0,1), \;\; \theta_{j1}|\theta_{(j-1)1} \sim Uniform(0, \theta_{(j-1)1}). \tag{4.18}$$

Similarly, we define the prior for $\boldsymbol{\theta}_2$ as

$$\pi(\boldsymbol{\theta}_2) = \pi(\theta_{12})\pi(\theta_{22}|\theta_{12}) \ldots \pi(\theta_{M2}|\theta_{(M-1)2})$$
$$\text{with } \theta_{12} \sim Uniform(0,2], \;\; \theta_{j2}|\theta_{(j-1)2} \sim Uniform(0, \theta_{(j-1)2}). \tag{4.19}$$

Let the set of hyper parameters in the process layer be denoted by $\xi_2 = (\beta_\mu, \theta_{1\mu}, \theta_{2\mu}, \theta_{1\rho}, \theta_{2\rho}, \theta_{1\alpha}, \theta_{2\alpha}, \gamma_1, \ldots, \gamma_n)$. Then conditional on $\xi_2$, the process layer model is given by:

$$\pi(\xi_1|\xi_2) \propto f(\boldsymbol{\mu}|\beta_\mu, \theta_{1\mu}, \theta_{2\mu}) \times f(\boldsymbol{\rho}|\theta_{1\rho}, \theta_{2\rho}) \times f(\boldsymbol{\alpha}|\theta_{1\alpha}, \theta_{2\alpha}) \times f(\Pi|\gamma_1, \ldots, \gamma_n).$$

*Priors*

We assume $\beta_\mu$ to have a Normal $(0,\Sigma_\mu I)$ distribution for some large $\Sigma_\mu$. The choice of priors for $(\theta_{1\mu}, \theta_{2\mu}, \theta_{1\rho}, \theta_{2\rho}, \theta_{1\alpha}, \theta_{2\alpha})$ remain same as in Section IV.1.4. We further assume that $\gamma_1, \ldots, \gamma_n$ are independent and identically distributed realizations from

an exponential distribution with mean 0.5. Thus, the prior model is given by

$$\pi(\xi_2) \propto f(\beta_\mu) \times \prod_{i=1}^{n} f(\gamma_i).$$

Then, the joint posterior distribution of $\xi = (\xi_1, \xi_2)$ is proportional to :

$$L_M(\mathbf{Y}|\xi_1) \times \pi(\xi_1|\xi_2) \times \pi(\xi_2). \tag{4.20}$$

### IV.3.2. **Prediction**

The posterior predictive density of the realizations $\mathbf{Y}_0^{k\times 1} = \mathbf{Y}(S_0) = (\mathbf{Y}(S_{01}), \ldots, \mathbf{Y}(S_{0k}))$ at unobserved locations $(S_{01}, \ldots, S_{0k})$ is given in the expression (4.9). We also assume that $\mathbf{Y}_0$ arise from the same random field as do the observations $\mathbf{Y}$. The $j^{th}$ component dispersion matrix for the augmented data vector $(\mathbf{Y}, \mathbf{Y}_0)^T$ is denoted by

$$\Sigma_j^{*(n+k\times n+k)} = \begin{pmatrix} \Sigma_j^{n\times n} & \Sigma_{12j} \\ \Sigma_{21j} & \Sigma_{22j}^{k\times k} \end{pmatrix}$$

. Further we set $u_i^* \stackrel{\Delta}{=} F_i(y_i; \mu_i, \rho_i, \alpha_i)$. Then from (4.17), we obtain the joint distribution of $(\mathbf{Y}, \mathbf{Y}_0|\boldsymbol{\eta}, \Pi)$ to be :

$$\begin{aligned} L_M(\mathbf{Y}_0, \mathbf{Y}|\boldsymbol{\eta}, \Pi) &\propto \sum_{j=1}^{M} \pi_j g_{0,\Sigma_j}\left(Q_g^{-1}(u_1^*), \ldots, Q_g^{-1}(u_{n+k}^*)\right) \\ &\times \prod_{i=1}^{n+k} \frac{f_i(y_i; \mu_i, \rho_i, \alpha_i)}{q_g\left(Q_g^{-1}(u_i^*)\right)}. \end{aligned} \tag{4.21}$$

Combining (4.17) and (4.21), we obtain the conditional distribution of $\mathbf{Y}_0$ given $\mathbf{Y}, \boldsymbol{\eta}$ and $\Pi$ as

$$f(\mathbf{Y}_0|\mathbf{Y}, \boldsymbol{\eta}, \Pi) \propto \sum_{j=1}^{M} \pi_j g_{0,\Sigma_j}\left(Q_g^{-1}(u_1^*), \ldots, Q_g^{-1}(u_{n+k}^*)\right) \prod_{i+1}^{n+k} \frac{f_i(y_i; \mu_i, \rho_i, \alpha_i)}{q_g\left(Q_g^{-1}(u_i^*)\right)}. \tag{4.22}$$

Once we have (4.22), we can approximate the posterior predictive density $f(\mathbf{Y}_0|\mathbf{Y})$ using the algorithm outlined in Section IV.1.5.

*Computation*

The MCMC algorithm to draw from the posterior distribution obtained in (4.20) is described below

1. Sample $\alpha$ from the target distribution $L_M(\mathbf{Y}|\boldsymbol{\eta}, \Pi)\pi(\alpha)$ using random walk proposal.

2. Sample $\boldsymbol{\beta}$ from the target distribution $L_M(\mathbf{Y}|\boldsymbol{\eta}, \Pi)\pi(\boldsymbol{\beta})$ using random walk proposal.

3. Sample $\rho$ using a Metropolis-Hastings scheme with Gamma(25,0.5) as a proposal distribution.

4. Sample $(\theta_{11}, \theta_{21}, \ldots, \theta_{M1})$ individually using the Metropolis scheme with Unif(0,1) as the proposal distribution.

5. Sample $(\theta_{12}, \theta_{22}, \ldots, \theta_{M2})$ individually using the Metropolis scheme with Unif(0,2] as the proposal distribution.

6. For model of order $M$, draw $\Pi = (\pi_1, \pi_2, \ldots \pi_M)$ using a Metropolis-Hastings scheme with Dirichlet as the proposal distribution.

We use posterior model probability as the model selection criterion. We consider up to models of order $K$ and these models are denoted by $M_1, M_2, \ldots, M_K$ respectively. Then the posterior model probability corresponding to $M_k$ is given by

$$P(M_k|\mathbf{Y}) = \frac{P(\mathbf{Y}|M_k)P(M_k)}{\sum_{r=1}^{K} P(\mathbf{Y}|M_r)P(M_r)} \tag{4.23}$$

Table 5.: Posterior probabilities for models of various order

| Copula | Marginals | Order | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Gaussian | Skew-normal | 0.0000 | 0.0001 | 0.0002 | **0.5554** | 0.4443 |
| | Skew-$t$ | 0.0495 | 0.0247 | 0.0124 | **0.5075** | 0.4060 |
| $T_{10}$ | Skew-normal | 0.0018 | 0.0009 | 0.0005 | **0.5538** | 0.4430 |
| | Skew-$t$ | 0.0845 | 0.0423 | 0.0211 | **0.4374** | 0.3787 |
| Logistic | Skew-normal | 0.0046 | 0.0023 | 0.0011 | **0.5511** | 0.4409 |
| | Skew-$t$ | 0.0889 | 0.0444 | 0.0222 | **0.4692** | 0.3753 |

where $P(M_k)$ is the prior probability associated with the model $M_k$ and $P(\mathbf{Y}|M_k) = \int L_{M_k}(\mathbf{Y}|\xi_{1k}, M_k)\pi(\xi_{1k}|\xi_{2k}, M_k)\pi(\xi_{2k}|M_k)d\xi_k$, $\xi_k = (\xi_{1k}, \xi_{2k})$ being the vector of the all parameters in the model $M_k$. Since $P(\mathbf{Y}|M_k)$ is not of closed form in this case, we use Monte-Carlo method we approximate the integral. Once we have samples from the posterior distribution of the parameters given in (4.20), we exploit them to estimate these probabilities.

### IV.3.3. **Simulation: non-stationary random field**

To study the ability of the proposed mixture model for identifying the true model even under complete misspecification, we perform an extensive simulation study. We generate six data sets considering the six models under our consideration (three copula each having two types of marginals) as the true model. Each data set contains 100 data points from a four component mixture model. We use six proposed models to analyze these data sets to check their robustness under model misspecification. We consider up to the fifth order model. The posterior model probabilities of each order are shown in Table 5.

It is clear that all the proposed models are able to identify the correct order. Subsequently, in Table 6., we compare the predictive accuracies of the all the fourth

Table 6.: Comparison among various mixture-copula models

| Fitted Model | **True** | | | **Model** | | |
|---|---|---|---|---|---|---|
| | Gaussian Skew-normal | Gaussian Skew-$t$ | $T_2$ Skew-normal | $T_2$ Skew-$t$ | Logistic Skew-normal | Logistic Skew-$t$ |
| Gaussian Skew-normal | 0.5644 (85%) | 1.2625 (71%) | 1.1140 (92%) | 2.3575 (69%) | 0.7214 (91%) | 2.48 (75%) |
| Gaussian Skew-$t$ | 0.5664 (89%) | 0.5149 (94%) | 1.1277 (92%) | 0.4851 (91%) | 0.7471 (93%) | 0.8727 (81%) |
| $T_2$ Skew-normal | 0.680 (88%) | 1.6281 (72%) | 0.833 (95%) | 1.2358 (84%) | 0.6682 (95%) | 1.4243 (80%) |
| $T_2$ Skew-$t$ | 0.7662 (94%) | 0.5565 (95%) | 0.8425 (95%) | 0.4211 (96%) | 0.7184 (96%) | 0.7090 (94%) |
| Logistic Skew-normal | 0.8637 (87%) | 1.2529 (71%) | 1.0598 (93%) | 1.6727 (73%) | 0.5381 (94%) | 1.3349 (80%) |
| Logistic Skew-$t$ | 0.8857 (92%) | 0.5717 (91%) | 1.0748 (95%) | 0.5046 (90%) | 0.5539 (94%) | 0.5925 (92%) |

order models.

Once again we notice that the MSPE of the models with correctly chosen marginals are lower than the ones with misspecified marginals, reinforcing our belief that the proper selection of marginals is of paramount importance. Also, it is to be noted that among all the misspecified models, the performance of the $t_2$-copula-skew $t$-marginals model is most stable. This indicates the robustness of the above mentioned model.

### IV.3.4. **Example: spatial permeability prediction**

Petroleum reservoirs are complex geological formations that exhibit a wide range of physical and chemical heterogeneities. These heterogeneities span over multiple length scales and are impossible to describe in a deterministic fashion. Geostatistics, and more specifically, stochastic modeling of reservoir heterogeneities are being increasingly considered by reservoir analysts and petroleum engineers for their poten-

tial in generating more accurate reservoir models together with realistic measures of spatial uncertainty. The goal of reservoir characterization is to provide a numerical model of reservoir attributes such as hydraulic conductivities (permeability), storativities (porosity) and fluid saturation. These attributes are then used as inputs into complex transfer functions represented by various flow simulators to forecast future reservoir performance and oil recovery potential. In predicting future reservoir performance, it is imperative to have a geological model that can be considered a 'plausible' replica of the actual reservoir with acceptable uncertainty. Towards this objective, we need more flexible modeling approaches to reproduce complex geological/morphological patterns and the wide variety of architectural heterogeneities observed in petroleum reservoirs.

In most flow situations, the single most influential input is the permeability spatial distribution. Permeability is an important concept in porous media flow (such as flow of underground oil). Physically, permeability arises both from the existence of pores and from the average structure of the connectivity of pores. Mathematically, fluid flow can typically be described by Darcy's law, which states that for steady-state flow in a porous medium, $v = -\rho\frac{\delta p}{L}\frac{1}{\mu}$, where $\rho$ is the permeability, $v$ is the volume flux per surface area of some region of length $L$, $\mu$ is the viscosity and $p$ is the pressure. The key role of permeability is evident from Darcy's law. In practice, therefore, dealing with the variability and uncertainties about permeability is critical for modeling porous flow.

Hence, permeability predictions are a vital aspect of a reservoir description but due to petrophysical variations rooted in digenesis, grain size variation, cementation, we observe highly heterogeneous behavior of the process at different regions of the reservoir (Lee et al. 1999). Modeling this heterogeneities is important as it has effect on the amount of recovered oil. Usually single stationary process fails to capture this

Table 7.: Posterior probabilities for models of various order

| Copula | Marginals | Order | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Gaussian | Skew-normal | 0 | 0 | 0.3117 | **0.4378** | 0.2505 |
| | Skew-$t$ | 0 | 0 | 0.0056 | **0.5525** | 0.4420 |
| $T_{10}$ | Skew-normal | 0 | 0.002 | 0.2071 | **0.4394** | 0.3515 |
| | Skew-$t$ | 0 | 0 | 0.1104 | **0.4942** | 0.3954 |
| Logistic | Skew-normal | 0.0042 | 0.1130 | 0.2221 | **0.3670** | 0.2936 |
| | Skew-$t$ | 0 | 0 | 0.18 | **0.4556** | 0.3645 |

huge heterogeneity and mixture models will be ideal to explore this data.

The Schneider Buda field is located in the Wood County, Texas. The field is discovered in recently. The field structure is an anticline, ten kilometers by eight kilometers, with the major axis N-S trending. Furthermore we will concentrate on permeability (Schon 1996; Barman et al. 1998; Lee et al. 1999) since it is the most important (and hard to determine) property for all reservoir problems - it controls whether the rock can deliver or transmit fluids or not. More precise information can be found at Peddibhotla et al. (1996). The scientists believe that the permeability field will be non-stationary due to presence of several barriers. The permeability is measured in 35 spatial locations and is expressed in the unit mD. The data is then centered around the observed mean and scaled by the observed standard deviation. This standardized data have support in $\mathbb{R}$ and thus allow us to assume the Skew Normal or Skew $t$ marginal processes.

Note the presence of skewness and multiple modes in the data. We fit a Gaussian copula model, a $t$ copula model and a logistic copula model each with Skew Normal and Skew $t$ marginals and obtain the Bayes Factor for each copula-marginal combination. We consider up to models of order five for each combination. Table 7. shows the posterior probabilities obtained for models of various orders. It is observed that

Table 8.: Comparison among various mixture-copula models

| Copula | Marginals | MSPE | 95% PI coverage |
|---|---|---|---|
| Gaussian | Skew-normal | 86.1247 | 91.43% |
| | Skew-$t$ | 58.2280 | 91.41% |
| $T_{10}$ | Skew-normal | 99.8497 | 92.29% |
| | Skew-$t$ | 50.2603 | 92.57% |
| Logistic | Skew-normal | 92.7845 | 91.43% |
| | Skew-$t$ | 71.3797 | 91.57% |

all the copula-marginal combinations choose the fourth order model. The predictive performances of the fourth order model corresponding to various copula marginal combinations are given Table 8..

Again, we notice that the $t_{10}$ copula model with Skew $t$ marginals has the best predictive performance among all the fourth order models. The choice of this copula degrees of freedom is based on exploratory analyses. Additionally, The Bayes factor in favor of this model as compared to a single component Gaussian process model is $1.69 \times 10^4$. It indicates an overwhelming support for the chosen non-stationary model. Once again, we believe that the shape parameter, $\boldsymbol{\alpha}$, to be the key parameter for inferential purpose. We plot the posterior median of $\boldsymbol{\alpha}$ for different sites along with the 95% credible interval in Figure 10(a). Almost none of the credible intervals contain 0 indicating a significant departure from Gaussianity. In Figure 10(b), we plot the kernel smoothed density of the observed data along with the posterior predictive density which shows satisfactory fit.

## IV.4. **Non-elliptical copulas for extreme observations**

### IV.4.1. **Extreme value processes**

In this section, we develop copula based spatial models for extreme observations. Suppose $(Y_{j1}, \ldots, Y_{jk})$ are independently and identically distributed random variable. Let

$$\begin{aligned} M_1 &= \text{Max}(Y_{11}, \ldots, Y_{1k}) \\ M_2 &= \text{Max}(Y_{21}, \ldots, Y_{2k}) \\ &\cdots \\ M_n &= \text{Max}(Y_{n1}, \ldots, Y_{nk}). \end{aligned}$$

Univariate extreme value theory suggests

$$\underset{k \to \infty}{\text{Lim}} \, \text{P}\left(\frac{M_i - b_i}{a_i} \leq y\right) = F(y)$$

for two sequences of real numbers $a_i > 0$ and $b_i$. If $F(y)$ is non-degenerate, it either belongs to the Gumbel, the Fréchet or the Weibull family of distribution, which can all be expressed under the class of generalized extreme value (GEV) distributions with density function given by:

$$f_{gev}(y; \mu, \rho, \alpha) = \frac{1}{\rho}\left(1 + \alpha\frac{y-\mu}{\rho}\right)^{-(1/\alpha+1)} e^{-\left(1+\alpha\frac{y-\mu}{\rho}\right)^{-1/\alpha}}$$

for $y : 1 + \alpha(y - \mu)/\rho > 0$ with $\mu \in \mathbb{R}$ being the location parameter, $\rho > 0$ being the scale parameter and $\alpha \in \mathbb{R}$ being the shape parameter. The value of $\alpha$ determines the subfamily with $\alpha = 0$ yields the Gumbel Distribution, $\alpha > 0$ corresponds to Fréchet distribution with heavy upper tails, while $\alpha < 0$ corresponds to Weibull distribution with bounded upper tails. Here, we wish to model the joint distribution

of $(M_1, \ldots, M_n)$. To derive this joint distribution, we make use of the multivariate extreme value copulas.

According to the theory of multivariate extreme value copulas, the joint distribution of $(M_1, \ldots, M_n)$ can be expressed, subject to the continuity conditions, as

$$F_{\boldsymbol{M}}(m_1, \ldots, m_n) = \mathbb{C}(F_{M_1}(m_1), \ldots, F_{M_n}(m_n)) = \mathbb{C}(u_1, \ldots, u_n)$$

where $\mathbb{C}$ satisfying the property $\mathbb{C}(U_1^t, \ldots, U_n^t) = \mathbb{C}^t(U_1, \ldots, U_n)$ for all $t > 0$. Pickands (1981), de Haan and Resnick( 1993), Hall and Tajvidi (2000) developed nonparametric methods to model the copula structure where as Tawn (1988), Coles and Tawn (1991,1994) adopted parametric approaches to model this structure. In Bayesian literature, the common practice is to use Bayesian hierarchical spatial models (Banerjee et al., 2004) where at the first stage the responses are assumed to be conditionally independent. This approach was taken by Cooley et al. (2007) and Sang and Gelfand (2009) in their analysis of extreme precipitation events. Although, Smith et al. (1997) jointly modeled the responses by using multivariate extreme value copula, the dependence structure was not unrestricted. Thus, none of these models enable us to explicitly incorporate the spatial dependence structure in the likelihood itself.

To alleviate these problems, we make use of a parametric multivariate extreme value copula which possesses a closed form distribution function and a flexible dependence structure can be incorporated there. We consider the family

$$\mathbb{C}(u_1, \ldots, u_n) = \exp\left[-\left\{\sum_{i=1}^{n} z_i^{\delta} - (n-1)^{-1} \sum_{1 \le i < j \le n} \left(z_i^{-\delta\Sigma_{ij}} + z_j^{-\delta\Sigma_{ij}}\right)^{-1/\Sigma_{ij}}\right\}^{1/\delta}\right] \tag{4.24}$$

where $z_i = -\log(u_i), u_i = F_{M_i}(m_i)$, $\Sigma_{ij}$ and $\delta \ge 1$ are the dependence parameters. While $\delta$ controls the global dependence, $\Sigma_{ij} = \theta_1^{l_{ij}^{\theta_2}}$ controls the pairwise dependence. Note that, $\Sigma_{ij}$ is a function of $l_{ij}$, which is the Euclidean distance between two loca-

tions. It also contains information about the range and the smoothness of the random field via the parameters $\theta_1$ and $\theta_2$, respectively. Thus the spatial information is imbibed in the joint distribution of $M_1, \cdots, M_n$ via $\Sigma_{ij}$. The role of these dependence parameters in achieving a wide range of dependence and their interplay is discussed in greater details in the following section. Note that the dependence parameters are defined on the $U$ process and not on the $M$ process. The family in (4.24) belongs to the class of multivariate extreme value distributions because the exponent is homogeneous of order 1 as a function of $z_1, \ldots, z_n$. Since, the class of multivariate extreme value distribution is essentially the class of max-stable distributions with non-degenerate marginals (Resnick, 1987), hence (4.24) belongs to class of max-stable distributions as well. This family is essentially a subfamily of the multivariate extreme value copulas introduced by Joe and Hu (1996).

The fact that we cannot obtain a multivariate extreme value distribution by combining univariate extreme value marginals with elliptical copulas prompted us to introduce the non-elliptical extreme value copula. Unlike, traditional extreme value copulas like Frank copula or Gumbel copula, the one considered here has unrestricted dependence structure. Moreover, (4.24) has closed form of distribution function which is an advantage over the multivariate extreme value distributions introduced by Joe (1994, 1996). Additionally, the unique feature of family (4.24) is that it is dimensionally consistent and hence gives rise to a valid random field which is necessary to model a spatial process.

### IV.4.2. **Properties**

*Dimensional consistency*

*Result 3.5:* The copula formulation in (4.24) is dimensionally consistent, in the sense that, if we integrate out $M_n$ (say) then the joint distribution of $M_1, \ldots, M_{n-1}$ is given

by

$$\mathbb{C}(u_1,\ldots,u_{n-1}) = \exp\left[-\left\{\sum_{i=1}^{n-1} z_i^{\delta} - (n-2)^{-1}\sum_{1\le i<j\le n-1}\left(z_i^{-\delta\Sigma_{ij}} + z_j^{-\delta\Sigma_{ij}}\right)^{-1/\Sigma_{ij}}\right\}^{1/\delta}\right] \tag{4.25}$$

*Proof.* See Appendix B □

*Tail dependence*

The coefficient of tail dependence is a scalar measure that relates to the behavior of the tails of a distribution. It is basically a summary of the extremal dependence inherent in a bivariate random vector and can be expressed in terms of copulas. Thus, in contrast to other dependence measures such as linear correlation they are not influenced by the marginal distributions of the random vector (Embrecht et al. (2001)). Since we are concerned with the upper extreme values, we concentrate on the behavior of the bivariate marginals in their upper tails only. This upper tail dependence measure quantifies the probability of one random variate being extreme, given that the other one is extreme too. From the definition, the pairwise upper tail dependence coefficient between $M_i$ and $M_j$ is given by

$$\lambda_{ij} = \underset{u\to 1-}{\text{Lim}} \quad \text{P}[M_i > F_{M_i}^{-1}(u) | M_j > F_{M_j}^{-1}(u)]$$

provided the limit $\lambda_{ij} \in [0,1]$ exists. If $\lambda_{ij} \in (0,1]$, $M_i$ and $M_j$ are said to be dependent in the upper tail in the class of MEV copulas, if $\lambda_{ij} = 0$, then they are asymptotically independent (Demarta, 2007).

For the multivariate extreme value copula described in (4.24), the bivariate marginals are given by

$$\mathbb{C}(u_i,u_j) = \exp\left[-\left\{z_i^{\delta} + z_j^{\delta} - \left(z_i^{-\delta\Sigma_{ij}} + z_j^{-\delta\Sigma_{ij}}\right)^{-1/\Sigma_{ij}}\right\}^{1/\delta}\right]$$

Following Joe and Hu (1996), the pairwise upper tail dependence coefficient derived from the above bivariate marginal is given by

$$\lambda_{ij} = 2 - \left[2 - 2^{-1/\Sigma_{ij}}\right]^{1/\delta} \tag{4.26}$$

which increases as $\Sigma_{ij}$ or $\delta$ increases.

Since we have claimed that the posited copula can handle wide range of dependence, it becomes necessary to discuss the behavior of the pairwise upper tail dependence coefficient obtained in (4.26). Given $\delta = 1$, as the distance between location $i$ and location $j$ increases, $\Sigma_{ij} \to 0$ and so does the upper tail dependence coefficient between $M_i$ and $M_j$, indicating that the extremal dependence for widely separated locations is virtually negligible. This makes intuitive sense. However, note that, if $\delta = 1$, then, even if $\Sigma_{ij} = 1$ (leading one to think of perfect dependence), the pairwise upper tail dependence coefficient cannot exceed 0.5, indicating a rather weak dependence. This situation is addressed by increasing $\delta$. In fact, for $\delta \approx 10$, and $\Sigma_{ij} = 1$, the pairwise upper tail dependence coefficient, $\lambda_{ij} \approx 0.99$. Thus, we see that $\delta$ plays the most crucial role in determining the extent of pairwise extremal dependence. An interesting point to be observed is that the pairwise upper tail dependence coefficient is bounded by 0 and 1 and is a function of the distance between two locations. Thus its behavior is analogous to an isotropic spatial correlation function.

*Posterior propriety*

In order to elicit information about $\delta$, we suggest a thorough exploratory study (see Chapter V for details). If the data empirically suggest weak pairwise extremal dependence, we fix $\delta$ to unity. If, however, the data reveal a strong pairwise extremal dependence, we specify a prior on $\delta$ such that the *a priori* mode is around 30. However, in absence of definitive idea about the extent of extremal dependence, one can

impose a noninformative prior of $\delta$. Under such specification, it can be shown that, under mild regularity conditions, the posterior is proper. The following result identifies the conditions under which the posterior propriety can be ascertained.

*Result 3.6:* Given the joint distribution function of $M_1, \cdots, M_n$ obtained in (4.24), define

$$A(\delta) = \left\{ \sum_{i=1}^{n} z_i^{\delta} - (n-1)^{-1} \sum_{1 \le i < j \le n} \left( z_i^{-\delta \Sigma_{ij}} + z_j^{-\delta \Sigma_{ij}} \right)^{-1/\Sigma_{ij}} \right\}^{1/\delta} \tag{4.27}$$

Then, an improper prior on $\delta$ yields a proper posterior under following conditions:

(i) $\int \frac{\tilde{\delta}^n}{\tilde{\delta} \prod_{i=1}^n m_i} F_{\boldsymbol{M}_t}(m_1, \cdots, m_n|\delta) d\delta = \frac{\tilde{\delta}^n}{\tilde{\delta} \prod_{i=1}^n m_i} \int F_{\boldsymbol{M}_t}(m_1, \cdots, m_n|\delta) d\delta$

(ii) $A(\delta)$ is at least twice differentiable and the MLE of $\delta$ exists.

(iii) $A''(\delta)|_{\delta=\delta_0} > 0$, where $\delta_0$ is the MLE of $\delta$ and $A'(\delta) = \frac{\tilde{\delta} A(\delta)}{\tilde{\delta} \delta}$ and $A''(\delta) = \frac{\tilde{\delta}^2 A(\delta)}{\tilde{\delta} \delta^2}$

where $\tilde{\delta}$ indicates the partial differential operator.

*Proof.* See Appendix A □

This ability to accommodate both strong and weak pairwise extremal dependence is an improvement over the usual Gaussian and $t$ copulas. For the Gaussian copula, $\lambda_{ij} = 0$ as long as $\Sigma_{ij} < 1$ and hence it excludes those class of models which show high pairwise extremally dependence. While for $t$ copula, $\lambda_{ij} > 0$ as long as $\Sigma_{ij} > -1$ and hence cannot be used to model observations that show weak extremal dependence.

### IV.4.3. **Hierarchical model**

*Choice of marginals*

Let $M_i = M(S_i)$ be the annual precipitation maxima observed at location $S_i, i = 1, \ldots, n$. The marginals for each $M_i$ is taken to be generalized extreme value distri-

bution with density function

$$f_{M_i}(m_i|\mu_i,\rho_i,\alpha_i) = \frac{1}{\rho_i}\left(1+\alpha_i\frac{m_i-\mu_i}{\rho_i}\right)^{-(1/\alpha_i+1)} e^{-\left(1+\alpha_i\frac{m_i-\mu_i}{\rho_i}\right)^{-1/\alpha_i}} \tag{4.28}$$

and distribution function

$$F_{M_i}(m_i|\mu_i,\rho_i,\alpha_i) = e^{-\left(1+\alpha_i\frac{m_i-\mu_i}{\rho_i}\right)^{-1/\alpha_i}} \tag{4.29}$$

with $m_i : 1+\alpha_i(m_i-\mu_i)/\rho_i > 0$. The choice of the marginal model is motivated by limiting distributions in extreme value theory. We now describe the hierarchical Bayesian model for component-wise maxima.

*Data model*

Using the multivariate extreme value model specified in (4.24), (4.28) and (4.29), and conditional on the parameters, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n), \boldsymbol{\rho} = (\rho_1,\ldots,\rho_n), \boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_n), \Sigma_\theta$, $\delta, \theta_1, \theta_2$ we get the joint density of $\boldsymbol{M} = (M_1,\ldots,M_n)$ as:

$$p(\boldsymbol{M}|\xi_1) \propto c(u_1,\ldots,u_n)\prod_{i=1}^{n} f_{M_i}(m_i|\mu_i,\rho_i,\alpha_i) \tag{4.30}$$

where $c(u_1,\ldots,u_n) = \frac{\delta^n \mathbb{C}(u_1,\ldots,u_n)}{\prod_{i=1}^n \delta u_i}$ and $\xi_1 = (\boldsymbol{\mu},\boldsymbol{\rho},\boldsymbol{\alpha},\delta,\theta_1,\theta_2)$. We obtain the derivative with help of symbolic computation software wherever possible.

*Process model*

In this layer we relate the parameters of the data layer to the covariates. Conditional on $\boldsymbol{\rho}$, we assume $\boldsymbol{\mu}|\boldsymbol{\rho} \sim N(X_i'\beta_\mu, diag(\boldsymbol{\rho}) \times \Sigma_{\theta_\mu})$. $X_i$ is the site-specific vector of covariates. We assume $\log(\boldsymbol{\rho}) \sim N(-0.5, \Sigma_{\theta_\rho})$. We assume a Uniform $(-c_\alpha, c_\alpha)$ prior on $\boldsymbol{\alpha}$. We further assume Uniform (0,1) distribution for $\theta_1$, Uniform (0,2] prior on $\theta_2$,

Uniform(1,50) for $\delta$. Then the process layer model is given by

$$p(\xi_1|\xi_2) \propto f(\boldsymbol{\mu}|X_i'\beta_\mu, diag(\boldsymbol{\rho}) \times \Sigma_{\theta_\mu}) \times f(\boldsymbol{\rho}| - 0.5, \Sigma_{\theta_\rho}) \tag{4.31}$$

where $\xi_2 = (\beta_\mu, \theta_\mu, \theta_\rho, c_\alpha)$

*Priors*

In this layer we assign priors for $\beta_\mu, \theta_\mu, \theta_\rho$ and $c_\alpha$. Since there is hardly any information about how the process model parameters are related to the covariates, we choose diffused priors for them. Thus, $\beta_\mu \sim N(0, c_\mu I)$ for some large $c_\mu$. Additionally, apriori $\theta_{1\mu}, \theta_{1\rho}$ are assumed to distributed independently as Uniform (0,1) while, $\theta_{2\mu}, \theta_{2\rho}$ are distributed independently as Uniform(0,2]. We further assume $c_\alpha \sim$ Uniform (0, 10). Then the joint prior distribution is given by

$$p(\xi_2) \propto N(0, c_\mu I). \tag{4.32}$$

Combining (5.4), (5.6) and (5.7), we get the joint posterior distribution of the parameters conditional on the data as

$$p(\xi|\boldsymbol{M}) \propto p(\boldsymbol{M}|\xi_1) \times p(\xi_1|\xi_2) \times p(\xi_2) \tag{4.33}$$

where $\xi = [\xi_1, \xi_2]$. We implement standard Metropolis within Gibbs sampler to draw samples from this joint posterior distribution.

### IV.4.4. **Example: extreme precipitation events across United States**

Estimation of chances of extreme precipitation events are important for flood planning purpose, which in turn is necessary for city planning, engineering and risk management. The National Weather Service (NWS) maintains a digital database that are primarily used to ascertain the chance of extreme precipitation at a particular location

(see *hdsc.nws.noaa.gov/hdsc/pfds/pfds-maps.html*). Estimation of such probabilities is difficult due to the necessity to interpolate at the locations where observations are not available. The principal objective of the proposed model is to interpolate over the study area and produce a map that can be used to calculate the chance of an extreme precipitation event at a particular location. As a default, we also produce a map of uncertainties associated with these predicted point estimates.

We illustrate our method by applying it on the monthly precipitation data collected over the continental United States in the year 1998. The dataset is gridded in $5^0 \times 5^0$ boxes yielding a total of 46 locations. Although we have monthly data available, we only model the annual maxima observed at each site, focusing exclusively on the block maxima approach to handle extreme observations. A natural alternative to our block maxima approach is the multivariate threshold approach with generalized Pareto marginals which will be explored in future.

We focus on three covariates: geographic coordinates, elevation and mean annual precipitation. For a non-homogeneous area with both mountain and plains, it is likely that elevation will have a significant influence on the events of extreme precipitation. It is also likely that mean annual precipitation will be a strong covariate. In fact, Coles and Tawn (1996) found that mean precipitation was a stronger covariate for extreme precipitation than elevation. Also note that, the mean precipitation data are highly correlated with the elevation data and take into account other factors such as slope, and meteorology. Cooley et al. (2007) suggested the use of different shape parameter for different orographic pattern. Here we use two values of $\boldsymbol{\alpha}$, one for the coastal region and the other for the inland regions each having an Unifrom $(-c_\alpha, c_\alpha)$ distribution and impose an Uniform (0,10) prior on $c_\alpha$. Just as earlier studies, the objective of this study is to perform efficient spatial interpolation. Hence, we assess the model adequacy using the same method described in Section IV.1.6. The mean square

prediction error comes out to be 56.58 while the empirical coverage probability of 95% prediction interval is 92.39%. The empirical coverage probability almost attains the nominal level further indicates that the estimates are not too optimistic in nature. The shape parameter $\boldsymbol{\alpha}$ is a key parameter that we need to draw inference on because the tail behavior of the marginal distribution depends exclusively on it. Figure 11(a) shows the median of the site-specific shape parameters. For most of the regions, $\alpha$ is significantly negative indicating a Weibull family. In Figure 11(b) we plot the kernel density estimates of the posterior distribution of $\boldsymbol{\alpha}$ for the inland and coastal stations. The posterior distributions are quite different for these two regions. The posterior median for $\boldsymbol{\alpha}$ corresponding to coastal regions (-1.62) is higher than that for the inland regions (-3.79). We observe bimodality in the distribution of the former with a second mode occurring at the positive part of its support. These two facts clearly indicate the prevalence of relatively heavier tails at the coastal region. The meteorological reason for the prevalence of such heavy tails in the coastal region is the seasonal development of tropical storms in the mid-Atlantic and Gulf of Mexico causing heavy precipitation at the eastern and south-eastern coast of USA. We have assumed the $\theta_1$ and $\theta_2$ to be *a priori* independent, however *a posteriori* they are dependent with correlation coefficient around 0.21. Figure 12(a) shows the heat map of predicted chance of extreme precipitation events across the entire study region. It is clear from this plot that most of the extreme precipitation events occur in the coastal region or in the great plains. There is not much evidence of heavy precipitation at mountainous locations. This finding agrees with that of Jarrett (1990, 1993) who claimed that the hydrologic and paleohydrologic evidence shows that intense rainfall does not occur at higher elevations. Figure 12(b) shows the uncertainty associated with this heat map. The levels of uncertainty are high in the desert locations where no stations are located, and in areas of very high elevation where the model is forced to extrapolate.

Figure 13 shows the kernel smoothed densities of the actual observations along with the kernel smoothed predictive density. It seems that the proposed model fits the data adequately.

## IV.5. **Concluding remarks**

In this chapter, we have proposed a class of novel spatial model which can accommodate the non-Gaussian nature of the data. These non-Gaussian models have been developed by the use of copulas which make them marginally consistent. We have extended our model in a mixture setup which can identify spatial clusters. We have generated multivariate distributions using elliptically contoured kernels with absolutely continuous marginals and have used them to model heavy tailed, skewed spatial processes. We have extended that methodology to develop a Generalized Linear Mixed Model for spatial data. We have also modeled extreme-valued spatial processes using this copula methodology. Thus, we have provided an alternative way of modeling multivariate extreme value processes with a flexible correlation structure. We have also circumvented the computational problem posed by the parametric families of multivariate extreme value distributions (Kotz and Nadarajah, 2000) which, with the exception of the time series logistic distributions ( Coles & Tawn, 1991), cannot handle a flexible correlation structure.

# CHAPTER V

## SPATIO-TEMPORAL MODELING OF EXTREMES: A CASE STUDY

In Chapter III and Chapter IV we have described how copulas can be used to model temporal and spatial dependence, respectively. A natural question that arises subsequently is: whether copulas can handle spatio-temporal dependence as well. In this chapter we strive to answer this very question. In particular, we concentrate on adopting copula methodology to handle the dependences exhibited by extremes observed over space and time. We aim to present and apply results from multivariate extreme value theory in terms of copulas.

Unlike the univariate case, the definition of an extreme event is not straightforward in a multivariate setup. The challenge lies in characterizing the extreme behavior of a random vector. For example, given a random vector of sufficiently high dimension, extreme behavior of a single component of it does not necessarily imply the extreme behavior of the whole vector. As a result, definition of multivariate extremes hinges on the probability that several marginals will give rise to extremes at the same time. In Chapter II we have shown that it is possible to separate the marginals and the behavior of random vectors using copulas. Following this approach we can apply well-known results from univariate EVT to the margins and subsequently concentrate on the dependence structure (i.e. on copulas). So, intuitively it appears that copulas will play a major role in the study of the multivariate extremes.

In this chapter we develop a highly flexible spatio-temporal model, in the Bayesian hierarchical paradigm, which provides a suite of predictive maps that can be used for inferring the probability of an extreme event at a particular location at a particular

time. However, due to limited temporal records and sparse spatial representation, such predictive maps are associated with non-trivial uncertainties. One advantage of our approach is that, we can quantify and, subsequently, produce a map of these predictive uncertainties in conjunction with the predictive maps of precipitation extremes.

*A precipitation atlas over time*

Hydro-meteorologists believe global warming and changing weather patterns are principle reasons behind a rising number of fatal flooding incidents. Various studies indicate that warming induced changes in the global water cycle could have more drastic impacts than ever imagined. As a consequence of this increase in global-mean precipitation, an increasing trend in extreme precipitation events have been observed over past few decades (Kunkel et al., 1999; Easterling et al., 2000). Continental US has witnessed a spate of heavy flooding in recent decades. In 1976 Big Thompson River flood killed 143 people in Colorado. In 1997, a flash flood in Antelope Canyon caused 11 casualties. The following year saw a flash flood occurring in San Marcos, Texas. More recently, flash floods wrecked havoc in Mount Rainier National Park in 2006 and in Davenport, Iowa in 2008. In 2009, the Red River Valley of eastern North Dakota and west-central Minnesota experienced a level flooding that occurred only once previously in the past 100 years. Although one might argue that such incidents are rare, understanding their frequency and intensity is essential for public safety (like flash flood warnings) and city planning (like urban drainage management). Engineers often need extreme precipitation statistics to device strategies for flood protection. To support these requirements, analyses and spatio-temporal prediction of extremal events are necessary. As a matter of fact the extreme value theory literature has grown considerably is the past few decades finding wide application is engineering,

oceanography and environmental sciences among others.

The goal of our application are the following: a) Develop yearly predictive atlases for annual precipitation maxima of the continental US. b) Quantification of the uncertainties associated with these predictions and produce atlases of the same. To this end, we use the Hulme dataset (Hulme, 1998) obtained from the University of East Anglia. The data consists of monthly precipitation estimates obtained for continental US from 1990 through 1998 for land gridboxes at a $5^0$ latitude by $5^0$ longitude resolution. This gridded dataset is derived from land gauge records that have been subjected to homogeneity procedures reported in Eischied et al. (1991) and Hulme (1992). Since data coverage varies over time, a 'quasi-fixed-grid' is defined which comprises of a subset of gridboxes that possess at least five years of data in every decade over the period 1900-1998. The result is a total of 46 grid cells across the continental US at each time point.

Literatures suggest statistically significant impact of geographical covariates on the latent spatial processes of the extremes (Cooley et al., 2007; Coles and Tawn, 1996). We focus on three such significant covariates, viz. the geographical coordinates, elevation and mean annual precipitation. For a non-homogeneous area with both mountain and plains, it is likely that elevation will have a significant influence on the events of extreme precipitation. It is also likely that mean annual precipitation will be a strong covariate. In fact, Coles and Tawn (1996) found that mean precipitation was a stronger covariate for extreme precipitation than elevation. Also note that, the mean precipitation data are highly correlated with the elevation data and take into account other factors such as slope, and meteorology.

*Spatial modeling of extremes*

There exists a rich depository of statistical literature concerned with modeling of extreme events. Coles (2001a) provides an excellent introduction and detailed discussion of various extant methodologies available to model such events. In particular, there exists considerable amount of studies focusing on development of models for extreme precipitation events. Coles and Tawn (1996a) used max-stable processes to model areal rainfall extremes. Durman et al. (2001) exploited global and regional climate models to simulate data and compare the occurrences of extreme daily precipitation events for present and future climates. Although a number of studies have used Bayesian methods to model extremes (Smith and Naylor,1987; Coles and Tawn, 1996b; Stephenson and Tawn, 2005), few models are available that efficiently incorporate information obtained from different spatial locations. Coles (2001b), Jagger and Elsner (2006) pooled data from various locations to increase the precision of the parameter estimates, but neither work attempted to model the spatial dependence explicitly. Casson and Coles (1999) develop a point process representation of exceedances over a threshold but did not extend their model in temporal domain. Cooley et al. (2007) developed a Bayesian hierarchical model that simultaneously model the spatial dependence and use that model to perform spatial interpolations. However, they also ignored the temporal dimension of the data. Huerta and Sanso (2007) and Sang and Gelfand (2009) are only exceptions that developed a complete hierarchical Bayesian spatio-temporal model using the dynamic linear model theory, but neither of them incorporated spatial and/or temporal information explicitly in the data model.

In a spatial set-up, the issues of asymptotic (in)dependence are always a concern in extreme value analyses. While asymptotically independent models are bound not to fit data that show asymptotic dependence, asymptotically dependent models

can be poor approximation for asymptotically independent variables, especially for finite samples (Ledford and Tawn; 1996, 1997). In the Bayesian paradigm, the most common technique to model extreme events, distributed over space, is to assume conditional independence at the data layer (Cooley et al., 2007, Huerta and Sanso, 2007). In order to incorporate dependence information at the data level, Sang and Gelfand (2009) combined univariate extreme value distribution with a Gaussian copula and developed a valid random field with unrestricted correlation structure. As another alternative model for such extremal events, multivariate extreme value (MEV) copula models were introduced (Tawn, 1988, Coles and Tawn, 1991, 1994, Ledford and Tawn, 1996). However, parametric models for such MEV copulas are very limited and do not allow unrestricted correlation structures. The dependence structure is obtained as by-product of these models. It is not easy to incorporate dependence information directly in the model.

However, while the first two models are theoretically inadequate to model dependences in extreme, the last model is restrictive in admitting correlation structures. A short discussion about the inadequacy of these approaches to model the dataset under consideration is provided in Section V.2.2. In order to circumvent these problems present in the extant models for extreme events, we propose a class of MEV distribution which is flexible enough to handle both weak and strong extremal dependence and at the same time allows unrestricted correlation. To this end, we appeal to the theory of MEV copulas. We allow the marginal processes to have their own unique distribution, with the mild restriction that the distribution function is invertible. Subsequently, we utilize a class of extreme value copula to combine these unique marginal processes to obtain a valid random field with explicit correlation structure. This maneuver enables us to incorporate the spatial dependence directly at the data layer of our model. We use two approaches to model the marginal processes, viz.,

the block-maxima approach and peak over threshold approach and compare their predictive performances using various diagnostic tools. We adopt a full Bayesian approach to make inference about the model parameters as well as to perform spatial interpolations and temporal predictions.

## V.1. **Exploratory analysis**

Since we are stressing on incorporating dependence in the data model itself, we need to ascertain whether the data actually reveal spatial dependence explicitly. So we run the Moran's $I$ randomization test for spatial dependence (Moran, 1950; Banerjee et al., 2004) on the annual maxima observed at 46 locations at each of the 99 time points, using the first order neighbor proximity. Figure 14 shows the plot of the $p-$value associated with the test at each time point. Note that, with a few exceptions, all the $p-$values are below the 5% level of significance, indicating a significant spatial dependence.

Now, we need to explore if we can parameterize the covariance structure. To this end we fit a binned empirical variogram and subsequently fit three parametric variograms exponential, Gaussian and spherical, to the data (Figure 15). As a diagnostic check, we obtain the mean square error (mse) to compare the empirical versus the fitted estimates. It turns out that the mse for exponential variogram is minimum (0.0013) followed by the mse of the spherical one (0.0022) whereas the Gaussian one has the maximum mse (0.0026). The fact that the exponential variogram performs better than the rest suggests that we should explicitly take into account this particular form of spatial dependence present among the marginal processes. In particular, we assume a generalized exponential covariance structure given by $\Sigma_{\boldsymbol{\theta}}(l) = \exp(-\nu l^{\theta_2}) = \theta_1^{l^{\theta_2}}$ where $l$ is the Euclidean distance between locations, $\nu > 0, \quad \theta_1 = \exp(-\nu)$ and $\boldsymbol{\theta} = [\theta_1, \theta_2]$.

To address the issue of asymptotic dependence, we resort to the $\chi$ and $\bar{\chi}$ diagnostic plots of Coles, Heffernan and Tawn (1999). Strong extremal dependence gives $\chi \in (0, 1]$ and $\bar{\chi} = 1$. Figures 16(a) and 16(b) shows the $\chi$ and $\bar{\chi}$ plots for location pair 3 and 18, while Figures 16(c) and 16(d) show the same for location pair 10 and 13. Note that $\chi \in (0, 1]$ and $\bar{\chi} \approx 1$ for both the pairs indicating evidence strong extremal dependence. Location pair 3 and 18 represent the strongest evidence of extremal dependence, while location pair 10 and 13 represent the weakest evidence of extremal dependence. Behavior of pairs not shown lies somewhere in between.

To study the temporal dependence, we plot the autocorrelations of the annual maxima time series observed at each location. Only a few of the locations in the eastern and the northern part of the continental U.S. reveal an underlying first order AR process. Most of the locations fail to show any significant temporal dependence. Figure 17 shows the autocorrelation plot of the above time series observed at four different climatological regions of the continental U.S.

Taking cognizance of the fact that the data show a marked spatial but feeble temporal dependence, we incorporate the spatial dependence directly in data model, and include the temporal information in the process layer by allowing the regression parameters to follow an AR(1) process.

## V.2. **Models**

### V.2.1. **The copula model**

Let $M_{i,t}$ denote the annual maximum precipitation at location $i$, $i = 1, \cdots, n$ at time $t$, $t = 1, \cdots, T$. Here we wish to obtain the joint distribution of $(M_{1,t}, \cdots, M_{n,t})$ at every time point $t$. To derive this joint distribution, we make use of the multivariate extreme value copulas.

Let $f_{M_{i,t}}(.)$ and $F_{M_{i,t}}(.)$ denote the distribution function and density function

of $M_{i,t}$ respectively. Then according to the theory of multivariate extreme value copulas, the joint distribution of $\boldsymbol{M}_t = (M_{1,t}, \ldots, M_{n,t})^T$ can be expressed, subject to the continuity conditions, as

$$F_{\boldsymbol{M}_i}(m_{1,t}, \ldots, m_{n,t}) = \mathbb{C}(F_{M_{1,t}}(m_{1,t}), \ldots, F_{M_{n,t}}(m_{n,t})) = \mathbb{C}(u_{1,t}, \ldots, u_{n,t})$$

where $\mathbb{C}$ satisfying the property $\mathbb{C}(U_{1,t}^k, \ldots, U_{n,t}^k) = \mathbb{C}^k(U_{1,t}, \ldots, U_{n,t})$ for all $k > 0$. Now consider the family

$$\mathbb{C}(u_{1,t}, \ldots, u_{n,t}) = \exp\left[-\left\{\sum_{i=1}^{n} z_{i,t}^{\delta} - (n-1)^{-1} \sum_{1 \le i < j \le n} \left(z_{i,t}^{-\delta\Sigma_{ij}} + z_{j,t}^{-\delta\Sigma_{ij}}\right)^{-1/\Sigma_{ij}}\right\}^{1/\delta}\right] \tag{5.1}$$

where $z_{i,t} = -\log(u_{i,t}), u_{i,t} = F_{M_{i,t}}(m_{i,t})$. As explained in Section IV.4.1, $\Sigma_\theta$ and $\delta \ge 1$ are the pairwise and global dependence parameters, respectively. Note that the dependence parameters are defined on the $U$ process and not on the $M$ process. The family in (5.1) belongs to the class of multivariate extreme value distributions because the exponent is homogeneous of order 1 as a function of $z_{1,t}, \ldots, z_{n,t}$. Since, the class of multivariate extreme value distribution is essentially the class of max-stable distributions with non-degenerate marginals (Resnick, 1987), hence (5.1) belongs to class of max-stable distributions as well. This family is essentially a subfamily of the multivariate extreme value copulas introduced by Joe and Hu (1996).

Unlike, traditional extreme value copulas like Frank copula or Gumbel copula, the one considered here has unrestricted dependence structure. Moreover, (5.1) has closed form of distribution function which is an advantage over the multivariate extreme value distributions introduced by Joe (1994, 1996). Additionally, the unique feature of family (5.1) is that it is dimensionally consistent and hence gives rise to a valid random field which is necessary to model a spatial process.

### V.2.2. **Existing models for extremes: a discussion**

In this section we discuss a few pitfalls present in the existing Bayesian approaches to model spatially distributed extreme observations.

*The conditional independence model*

In the Bayesian paradigm, the most common technique to model extreme events, distributed over space, is to assume conditional independence at the data layer (Cooley et al., 2007, Sang and Gelfand, 2009). The following result identifies the theoretical inadequacy of such model.

*Result 4.1:* If the joint distribution of $(M_{1,t}, \cdots, M_{n,t})$ is given by $\prod_{i=1}^{n} f_{M_{i,t}}(m_{i,t}|\xi)$, where $\xi$ is the vector of all the parameters, then the pairwise upper tail dependence coefficient between $M_{i,t}$ and $M_{j,t}$ is 0 under proper prior specification on $\xi$.

*Proof.* See Appendix A □

In other words, assumption of conditional independence at the data layer, that is, failure to incorporate the dependence information at the data model leads to independence in the upper tail. An immediate implication of the above result is that, the traditionally used conditionally independent data model cannot explain the dependence of very rare events at two specified sites, no matter how close they are.

*Gaussian copula based extreme value models*

The Gaussian copula based MEV model proposed by Sang and Gelfand (2009) also leads to independence in the upper tail and hence proves to be inadequate to handle variables showing strong extremal dependence. A proof of this can be obtained in Demarta (2007). Another problem with using Gaussian copula with GEV marginals is that, the resultant model does not belong to the family of Multivariate Extreme

Value distributions. Consequently, the properties of MEV distributions do not hold true for these models and hence cannot be used in the inferential purposes.

*t copula based extreme value models*

Instead of combining the GEV marginals using the Gaussian copula, one can use $t$ copula for the same purpose. One advantage of $t$ copula is that, it always leads to upper tail dependence (Demarta, 2007). As a result, this model excludes very weak extremally independent variables. However, in its most common parameterization, the $t$ copula model does not belong to the domain of attraction of the MEV distributions and hence would not be a suitable model for multivariate extremes. But we note, from Section V.1, that the present data do exhibit strong pairwise extremal dependence. So, we use a $t$ copula based extreme value model as our baseline model.

The MEV copula model presented in (5.1) alleviates all these problems. Since the exponent is homogeneous of order 1, it belongs to class of MEV distributions. Hence, it should prove theoretically adequate to model multivariate extremes. Already, in Section IV.4.2, we have shown that this model has the ability to handle both strong and weak extremal dependence. Furthermore, this model allows us to incorporate the spatial dependence directly at the data layer by specifying the $\Sigma'_{ij}$s.

### V.2.3. **The block maxima approach**

Let $Y_1, Y_2, \cdots$ be a sequence of iid random variables with unspecified distribution. For a given $n$, define $M_n = \text{Max}(Y_1, Y_2, \cdots, Y_n)$. Then the extreme value theory suggests

$$n \overset{\text{Lim}}{\to} \infty \, \text{P}\left(\frac{M_n - b_n}{a_n} \leq y\right) = F(y)$$

for two sequences of real numbers $a_n > 0$ and $b_n$. If $F(y)$ is non-degenerate, it either belongs to the Gumbel, the Fréchet or the Weibull family of distribution, which can

all be expressed under the class of generalized extreme value (GEV) distributions with density function and distribution function respectively given by:

$$f^{gev}_{M_{i,t}}(m_{i,t}|\mu,\rho,\alpha) = \frac{1}{\rho}\left(1+\alpha\frac{m_{i,t}-\mu}{\rho}\right)^{-(1/\alpha+1)} \exp\left[-\left(1+\alpha\frac{m_{i,t}-\mu}{\rho}\right)^{-1/\alpha}\right] \quad (5.2)$$

$$F^{gev}_{M_{i,t}}(m_{i,t}|\mu,\rho,\alpha) = \exp\left[-\left(1+\alpha\frac{m_{i,t}-\mu}{\rho}\right)^{-1/\alpha}\right] \quad (5.3)$$

for $m_{i,t} : 1+\alpha(m_{i,t}-\mu)/\rho > 0$ with $\mu \in \mathbb{R}$ being the location parameter, $\rho > 0$ being the scale parameter and $\alpha \in \mathbb{R}$ being the shape parameter. The value of $\alpha$ determines the subfamily with $\alpha = 0$ which yields the Gumbel Distribution, $\alpha > 0$ corresponds to Fréchet distribution with heavy upper tails, while $\alpha < 0$ corresponds to Weibull distribution with bounded upper tails.

In this approach we assign monthly precipitation for each location into annual blocks and allow the marginal distribution of each maximum to follow its own unique GEV. Given the time, we combine GEV marginals across space using the extreme value copula described in (5.1). Then conditional on $\mu, \rho, \alpha$, we assume the maxima to be independent across the years. Annual maxima at a location occur with a sufficient time lag between them justifying the assumption of conditional temporal independence. We now describe the hierarchical Bayesian model for component-wise maxima.

*Data model*

We assume that marginally $M_{i,t} \sim$ GEV $(\mathbf{X}_i\boldsymbol{\beta}_t, \rho_i, \alpha_i)$, where $\mathbf{X}_i$ is a $p$ dimensional vector of covariates associated with the location $i$ , $\boldsymbol{\beta}_t = [\beta_{1t}, \ldots, \beta_{pt}]^T$ is the set of time dependent regression parameters. We assume a second order stationary spatial process with isotropic generalized exponential covariance structure described earlier. For computational simplicity we assume the scale and the shape parameter to be spa-

tially dependent but temporally invariant, although it is conceptually straightforward to deal with a space-time dependent $\rho$ and $\alpha$. Then using the multivariate extreme value model specified in (5.1), (5.2) and (5.3) and assuming conditional temporal independence, we write the full likelihood for $\boldsymbol{M} = [\boldsymbol{M}_1, \cdots, \boldsymbol{M}_T]^T$ as

$$P_1^{gev}(\boldsymbol{M}|\xi_1) \propto \prod_{t=1}^{T} c(u_{1,t}, \ldots, u_{n,t}) \prod_{i=1}^{n} f_{M_{i,t}}^{gev}(m_{i,t}|\mathbf{X}_i\boldsymbol{\beta}_t, \rho_i, \alpha_i) \tag{5.4}$$

where $c(u_{1,t}, \ldots, u_{n,t}) = \frac{\delta^n \mathbb{C}(u_{1,t},\ldots,u_{n,t})}{\prod_{i=1}^{n} \delta u_{i,t}}$ and $\xi_1 = (\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\alpha}, \delta, \theta_1, \theta_2)$ with $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n), \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T]^T$. We obtain the derivative with help of symbolic computation software wherever possible.

*Process model*

In the second layer of our model, we construct a hierarchical structure that relates the parameters of the data layer to latent temporal process. We model the time dependent regression parameter as follows:

$\beta_{j,t} = \gamma_j \beta_{j,t-1} + \epsilon_t, \quad j = 1, 2, \cdots p$ where $\epsilon_t \sim N(0, \sigma_\beta^2)$. Assuming conditional independence of $\boldsymbol{\beta}_t$ given the AR parameters $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_p]^T$, we get the joint distribution of $\boldsymbol{\beta}$ as

$$\pi_{\boldsymbol{\beta}}^{gev}(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma_\beta^2) \propto \prod_{j=1}^{p} \prod_{t=1}^{T} \frac{1}{\sigma_\beta} \exp\left[-\frac{1}{2\sigma_\beta^2} \left(\beta_{j,t} - \gamma_j \beta_{j,t-1}\right)^2\right]. \tag{5.5}$$

Note that, in expression (5.5), we need to estimate the initial condition $\boldsymbol{\beta}_0$. For this purpose, we fit a purely spatial model at time point $t = 0$ and obtain its ML estimate, say $\tilde{\boldsymbol{\beta}}_0$. Subsequently, we assume $\tilde{\boldsymbol{\beta}}_0$ to be the known initial condition. Thus we have

$$\tilde{\boldsymbol{\beta}}_0 = \underset{\boldsymbol{\beta}_0}{\text{argmax}}\ c(u_{1,0}, \ldots, u_{n,0}) \prod_{i=1}^{n} f_{M_{i,0}}^{gev}(m_{i,0}|\mathbf{X}_i\boldsymbol{\beta}_0, \rho_i, \alpha_i)$$

We assume $\log(\boldsymbol{\rho}) \sim N(-0.5, \Sigma_{\theta_\rho})$, where $\Sigma_{\theta_\rho} = {\theta_{1\rho}}^{l^{\theta_{2\rho}}}$. We assume two values for the shape parameter $\boldsymbol{\alpha}$, one for the coastal stations and one for inland stations. With $\boldsymbol{\alpha}_{coastal} \sim$ Uniform $(-\alpha_c, \alpha_c)$ and $\boldsymbol{\alpha}_{inland} \sim$ Uniform $(-\alpha_m, \alpha_m)$. We further assume Uniform (0,1) distribution for $\theta_1$, Uniform (0,2] prior on $\theta_2$ and a Uniform prior on $\delta$.

Assuming *a priori* independence of the parameters in this layer we get the process layer model as

$$P_2^{gev}(\xi_1|\xi_2) \propto \pi_{\boldsymbol{\beta}}^{gev}(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma_\beta^2) \times \pi(\boldsymbol{\rho}| - 0.5, \Sigma_{\theta_\rho}) \tag{5.6}$$

where $\xi_2 = (\boldsymbol{\gamma}, \sigma_\beta^2, \theta_{1\rho}, \theta_{2\rho}, \alpha_c, \alpha_m)$.

*Priors*

Finally, we arrive at the last stage of the hierarchy where we assign priors on $\boldsymbol{\gamma}, \sigma_\beta^2, \theta_{1\rho}, \theta_{2\rho}$ and $c_\alpha$. We assume that $\gamma_j|\sigma_\gamma^2 \sim N(0, \sigma_\gamma^2)$ and $(\sigma_\gamma^2, \sigma_\beta^2)$ are independently distributed as Inverse Gamma(0.1,100). Additionally, *a priori* $\theta_{1\rho}$ and $\theta_{2\rho}$ are assumed to be distributed independently as Uniform (0,1) and Uniform(0,2] respectively. We further assume $(\alpha_c, \alpha_m)$ are independently distributed as Uniform (0, 10). Assuming independence of $\gamma_j$'s conditional on $\sigma_\gamma^2$, we get the joint prior distribution as

$$P_3^{gev}(\xi_2) \propto \prod_{j=1}^{p} \pi_{\gamma_j}(\gamma_j|\sigma_\gamma^2) \times \pi_{\sigma_\gamma^2}(\sigma_\gamma^2) \times \pi_{\sigma_\beta^2}(\sigma_\beta^2). \tag{5.7}$$

Combining the data model, the process model and the priors as obtained in (5.6) and (5.7), we get the joint posterior distribution of the parameters conditional on the data as

$$P^{gev}(\xi|\boldsymbol{M}) \propto P_1^{gev}(\boldsymbol{M}|\xi_1) \times P_2^{gev}(\xi_1|\xi_2) \times P_3^{gev}(\xi_2) \tag{5.8}$$

where $\xi = [\xi_1, \xi_2]^T$.

### V.2.4. **Peak over threshold approach**

Instead of directly modeling the annual maxima using a GEV distribution, the *peak over threshold* (POT) approach (North, 1980) models the exceedances over a high threshold value. It consists of three components:

(i) Determination of the threshold value, $\eta$.

(ii) Number of occurrences of exceedance over the threshold value, $\eta$, over a given period of time which is assumed to be governed by a Poisson process.

(iii) The excess values, i.e, the amount by which the threshold is exceeded have a Generalized Pareto distribution (GPD).

*Threshold selection*

We use the procedure of threshold selection similar to the one used by Kunkel et al. (1999). The outline of the algorithm is given below:

(i) The monthly precipitation time series corresponding to every location is ranked in descending order.

(ii) The amount of precipitation during the $N^{th}$ ranked event of a particular time series is considered as the threshold value for a specified return period ($r$) for that particular time series, where

$$N = \left[\frac{\text{Number of years of data to be analyzed}}{r}\right]$$

where [.] is the greatest integer operator. In this paper the numerator is 99 years and we choose a return period ($r$) of one year.

(iii) We end up with different threshold values for different time series. We choose the minimum of them as the threshold value for all locations.

The main assumption of this Poisson-GP approach is that the underlying process governing the exceedances over the threshold is IID (Leadbetter et al., 1983). In the present study, it is observed that for a given location, the exceedances tend to occur in clusters. To deal with this temporal dependence, we resort to the declustering technique (NERC, 1975; North, 1980). Thus if a location records exceedances over consecutive time points, we 'decluster' the data by keeping only the highest measurement of that cluster. Selecting the cluster maxima for each cluster makes the data approximately independent and thereby amenable to the Poisson-GP model.

After declustering, we assume that the number of clusters containing the exceedances over the time period follows a Poisson process (Leadbetter et al., 1983). We are now in a position to describe the formulation in a greater detail.

*Poisson model for cluster occurrence*

At the $i$th location, let $N_i$ be the number of clusters containing the events of exceedances over $\eta$ in the $t^{th}$ year. We assume $N_i \sim$ Poisson $(\lambda_i)$, where $\lambda_i$ is the mean number of clusters occurring per year for location $i$.

*GP model for exceedances*

Conditional on $N_i \geq 1$, the amounts by which the cluster maxima of location $i$ in the $t^{th}$ year, $M_{i_1,t}, \ldots, M_{i_{N_i},t}$, exceed the threshold are identically and independently distributed as GP$(\mu_{i,t}, \rho_i, \alpha_i)$, with distribution function given by

$$\mathrm{P}(M_{i_j,t} - \eta \leq m | M_{i_j,t} > \eta) = G(m|\mu_{i,t}, \rho_i, \alpha_i) = 1 - \left(1 + \alpha_i \frac{m - \mu_{i,t}}{\rho_i}\right)^{-1/\alpha_i} \quad (5.9)$$

for $m \geq \mu_{i,t}$ when $\alpha_i \geq 0$ and $m \leq \mu_{i,t} - \rho_i/\alpha_i$ when $\alpha_i < 0$. Now, we derive the marginal distribution of annual maxima for location $i$ as follows:

$$\begin{aligned} \text{Let } M_{i,t} &= \underset{1\leq j\leq N_i}{\text{Max}} \quad M_{i_j,t} \\ \mathrm{P}(M_{i,t} \leq m | M_{i,t} > \eta) &= \mathrm{P}(N_i = 0) + \sum_{n_i=1}^{\infty} \mathrm{P}(N_i = i, M_{i_1} \leq m, \ldots, M_{i_{n_i}} \leq m) \\ &= \exp\left[-\lambda_i \left(1 + \alpha_i \frac{m - \eta - \mu_{i,t}}{\rho_i}\right)^{-1/\alpha_i}\right] \end{aligned} \tag{5.10}$$

Differentiating (5.10) with respect to $M$ we get the marginal density of annual maxima at location $i$ as

$$\begin{aligned} f^{gpd}(m_{i,t}; \mu_{i,t}, \rho_i, \alpha_i) &= \frac{\lambda_i}{\rho_i}\left(1 + \alpha_i \frac{m_{i,t} - \eta - \mu_{i,t}}{\rho_i}\right)^{-1/\alpha_i - 1} \\ &\quad \exp\left[-\lambda_i \left(1 + \alpha_i \frac{m_{i,t} - u - \mu_{i,t}}{\rho_i}\right)^{-1/\alpha_i}\right] \end{aligned} \tag{5.11}$$

for $m_{i,t} - \eta \geq \mu_{i,t}$ when $\alpha_i \geq 0$ and $m_{i,t} - \eta \leq \mu_{i,t} - \rho_i/\alpha_i$ when $\alpha_i < 0$.

*Data model*

After obtaining the marginal density of the annual maxima for the POT approach, we employ the extreme value copula derived in (5.1) to combine the marginal densities of annual maxima to get the data layer model for a particular year. Then, assuming conditional temporal independence we formulate the full data layer model as:

$$P_1^{gpd}(\boldsymbol{M}|\xi_1) \propto \prod_{t=1}^{T} c(u_{1,t}, \ldots, u_{n,t}) \prod_{i=1}^{n} f_{M_{i,t}}^{gpd}(m_{i,t}|\mathbf{X}_i\boldsymbol{\beta}_t, \rho_i, \alpha_i) \tag{5.12}$$

where $U_{i,t} = F_{M_{i,t}}^{gpd}(m_{i,t}|\mathbf{X}_i\boldsymbol{\beta}_t, \rho_i, \alpha_i)$ is derived in (5.10) and $f_{M_{i,t}}^{gpd}(m_{i,t}|\mathbf{X}_i\boldsymbol{\beta}_t, \rho_i, \alpha_i)$ is derived in (5.11) and $\xi_1 = [\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\alpha}, \theta_1, \theta_2, \boldsymbol{\lambda}]^T$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)$.

*Process model*

We model the time dependent regression parameter as follows:

$\beta_{j,t} = \gamma_j \beta_{j,t-1} + \epsilon_t, \quad j = 1, 2, \cdots p$ where $\epsilon_t \sim N(0, \sigma_\beta^2)$. Assuming conditional independence of $\boldsymbol{\beta}_t$ given the AR parameters $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_p]^T$, we get the joint distribution of $\boldsymbol{\beta}$ as

$$\pi_{\boldsymbol{\beta}}^{gpd}(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma_\beta^2) \propto \prod_{j=1}^{p} \prod_{t=1}^{T} \frac{1}{\sigma_\beta} \exp\left[ -\frac{1}{2\sigma_\beta^2} \left(\beta_{j,t} - \gamma_j \beta_{j,t-1}\right)^2 \right]. \tag{5.13}$$

Once again, we estimate the initial condition $\boldsymbol{\beta}_0$ in the same way as outlined in Section V.2.3.

We assume $\log(\boldsymbol{\rho}) \sim N(-0.5, \Sigma_{\theta_\rho})$, where $\Sigma_{\theta_\rho} = {\theta_{1\rho}}^{l^{\theta_{2\rho}}}$. We assume two values for the shape parameter $\boldsymbol{\alpha}$, one for the coastal stations and one for inland stations. With $\boldsymbol{\alpha}_{coastal} \sim$ Uniform $(-\alpha_c, \alpha_c)$ and $\boldsymbol{\alpha}_{inland} \sim$ Uniform $(-\alpha_m, \alpha_m)$. We further assume Uniform (0,1) distribution for $\theta_1$, Uniform (0,2] prior on $\theta_2$ and a Uniform prior for $\delta$. We further assume $\lambda_i \sim$ Gamma$(\bar{n_i}, 1)$ where $\bar{n_i}$ is the observed mean number of clusters per year for location $i$. Assuming *a priori* independence, we get the joint distribution of $\boldsymbol{\lambda}$ as $\pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}) \propto \prod_{i=1}^{n} \pi_{\lambda_i}(\lambda_i)$

Assuming *a priori* independence of the parameters in this layer, we get the process layer model as

$$P_2^{gpd}(\xi_1|\xi_2) \propto \pi_{\boldsymbol{\beta}}^{gpd}(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma_\beta^2) \times \pi(\boldsymbol{\rho}| - 0.5, \Sigma_{\theta_\rho}) \times \pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}) \tag{5.14}$$

where $\xi_2 = [(\boldsymbol{\gamma}, \sigma_\beta^2, \theta_{1\rho}, \theta_{2\rho}, \alpha_c, \alpha_m)]^T$.

*Priors*

We assume $\gamma_j|\sigma_\gamma^2 \sim N(0, \sigma_\gamma^2)$ and $(\sigma_\gamma^2, \sigma_\beta^2)$ are independently distributed as Inverse Gamma (0.1,100). Additionally, *a priori* $\theta_{1\rho}$ and $\theta_{2\rho}$ are assumed to distributed

independently as Uniform (0,1) and Uniform(0,2] respectively. We further assume $(\alpha_c, \alpha_m)$ are independently distributed as Uniform (0, 10). Assuming conditional independence of $\gamma_j$'s given $\sigma^2_\gamma$, we get the joint prior distribution as

$$P_3^{gpd}(\xi_2) \propto \prod_{j=1}^{p} \pi_{\gamma_j}(\gamma_j|\sigma^2_\gamma) \times \pi_{\sigma^2_\gamma}(\sigma^2_\gamma) \times \pi_{\sigma^2_\beta}(\sigma^2_\beta). \tag{5.15}$$

Combining the data model, the process model and the priors as obtained in (5.12), (5.14) and (5.15) we get the joint posterior distribution as

$$P^{gpd}(\xi|\boldsymbol{M}) \propto P_1^{gpd}(\boldsymbol{M}|\xi_1) \times P_2^{gpd}(\xi_1|\xi_2) \times P_3^{gpd}(\xi_2) \tag{5.16}$$

where $\xi = [\xi_1, \xi_2]^T$.

We implement standard Metropolis within Gibbs sampler to draw samples from this joint posterior distributions described in (5.8) and (5.16).

Once the posterior samples are obtained, prediction for the $T + 1^{th}$ time point at an unobserved site $S_0$ can be obtained in two steps.

(i) Precipitation for the $T + 1^{th}$ year can be obtained by drawing samples from the target distribution

$$P(\boldsymbol{M}_{T+1}|\boldsymbol{M}_1, \ldots, \boldsymbol{M}_T) = \int_\xi P(\boldsymbol{M}_{T+1}|\xi) \times P(\xi|\boldsymbol{M}) d\xi \tag{5.17}$$

where $\boldsymbol{M}_t = [M_{1,t}, \ldots, M_{n,t}]^T$, $P(M_{T+1}|\xi)$ is the data layer model for the new observation similar to (5.4) or (5.12) and $P(\xi|\boldsymbol{M})$ is the posterior distribution obtained in (5.8) or (5.16)

(ii) Once the temporal prediction is obtained, we fix the time point and perform a spatial interpolation at an unobserved location $S_0$ by drawing samples from the

target distribution

$$P(M_{S_0}|M_{1,T+1},\ldots,M_{S,T+1}) = \int_{\xi_{T+1}} P(M_{S_0,T+1}|\xi_{T+1}) \times P(\xi_{T+1}|\boldsymbol{M}_{T+1}) d\xi_{T+1} \tag{5.18}$$

where $\xi_{T+1}$ is the set of parameters for the $T+1^{th}$ time point and $P(\xi_{T+1}|\boldsymbol{M}_{T+1})$ is its posterior distribution.

### V.2.5. $t$ copula model

We use a $t$ copula model for extreme values as a baseline model to compare the performance of the model proposed in (5.1). The logic behind choosing it as a baseline model instead of a Gaussian copula model is that, the latter leads to asymptotic independence and hence unsuitable for the present data under consideration which show strong asymptotic dependence.

The data layer model is given by

$$\begin{aligned} P_1^t(\boldsymbol{M}|\xi_1) \quad \propto \quad & \prod_{t=1}^{T} t_{\Sigma,k}\left(T_k^{-1}(F_{M_{1,t}}(m_{1,t};\mathbf{X}_1\boldsymbol{\beta}_t,\rho_1,\alpha_1)),\cdots,T_k^{-1}(F_{M_{n,t}}(m_{n,t},\mathbf{X}_n\boldsymbol{\beta}_t,\rho_1,\alpha_1))\right) \\ & \times \prod_{i=1}^{n} \frac{f_{M_{i,t}}(m_{i,t};\mathbf{X}_i\boldsymbol{\beta}_t,\rho_i\alpha_i)}{t_k(T_k^{-1}(F_{M_{i,t}}(m_{i,t};\mathbf{X}_i\boldsymbol{\beta}_t,\rho_i,\alpha_i)))} \end{aligned}$$

where $t_{\Sigma,k}$ denote the p.d.f of an $n$-variate $t$-distribution with covariance matrix $\Sigma$ and $k$ degrees of freedom. $T_k$ and $t_k$ denote the distribution function and density function of an univariate $t$-distribution with d.f $k$ and variance 1. For the block-maxima approach we use the expression obtained in (5.2) and (5.3) for $f_M(.;\mathbf{X}_i\boldsymbol{\beta}_t,\rho_i,\alpha_i)$ and $F_M(.;\mathbf{X}_i\boldsymbol{\beta}_t,\rho,\alpha_i)$ respectively and model the process and the prior layers are in similar fashion as in Section V.2.3 and V.2.3. For Poisson-GP approach we use expression obtained in (5.10) and (5.11) as the marginal cdf and pdf respectively and the process and prior layers analogous to the model described in Section V.2.4 and V.2.4. Based on exploratory analysis, we fix the degrees of freedom, $k$, of the $t$ copula to be 5.

However, one can make it more flexible by putting a Poisson prior on it.

## V.3. **Model evaluation**

*Posterior predictive model checks*

The goal is to ascertain whether the observed data are representative of the type of data we might expect under the model. We can assess the fit using draws from the posterior predictive distribution obtained in (5.17) and (5.18) to represent what we can expect under the model. Let $\boldsymbol{M}^{rep}$ denote a replication of the data with the same, but unknown, values of the parameters that produced the data $\boldsymbol{M}$. To assess the fit of the model we introduce a discrepancy measure $T(\boldsymbol{M}, \xi)$ to measure the overall fit of the model to the data. In this study, we use the usual chi-squared goodness-of-fit measure, that is

$$T(\boldsymbol{M}, \xi) = \sum_i \frac{(M_i - E(M_i|\xi))^2}{Var(M_i|\xi)}$$

Note that $T$ is a function of the parameters and hence has a posterior distribution. The fit of the model with respect to the discrepancy measure $T$ is assessed by comparing the posterior distribution of $T(\boldsymbol{M}, \xi)$ to that of $T(\boldsymbol{M}^{rep}, \xi)$. The comparison is carried out via simulation. We draw $N$ simulations $\xi^1, \xi^2, \cdots, \xi^N$ from the posterior distribution of $\xi$ given in (5.8) or (5.16) and then draw one $\boldsymbol{M}^{rep}$ from the predictive distribution, given in (5.17) and (5.18), using each simulated $\xi$. Thus we have $N$ draws from the joint posterior distribution $P(\boldsymbol{M}^{rep}, \xi|\boldsymbol{M})$. Then we compare the values of the realized discrepancy $T(\boldsymbol{M}, \xi^k)$ and the replicated discrepancy measures $T(\boldsymbol{M}^{rep,k}, \xi^k)$, $k = 1, 2, \cdots, N$. One way of comparing their joint posterior distribution is by plotting the pairs $(T(\boldsymbol{M}, \xi^k), T(\boldsymbol{M}^{rep,k}, \xi^k))$ in a scatterplot. If the points are far removed from a $45^0$ line, then the data generated by the model do not resemble the observed data as regards the measure $T$. Using these simulated values

we also obtain the posterior predictive $p-$value

$$\begin{aligned} p_p &= \mathrm{P}(T(\boldsymbol{M}^{rep},\xi) \geq T(\boldsymbol{M},\xi)|\boldsymbol{M}) \\ &= \int\int I_{\left[T(\boldsymbol{M}^{rep},\xi)\geq T(\boldsymbol{M},\xi)\right]} P(\boldsymbol{M}^{rep}|\xi)P(\xi|\boldsymbol{M})d\boldsymbol{M}^{rep}d\xi \end{aligned}$$

where $I_{[A]}$ is the indicator function for the event $A$. The $p-$value is estimated from the simulations as the proportion of the $N$ replications for which $T\left(\boldsymbol{M}^{rep},\xi\right) \geq T\left(\boldsymbol{M},\xi\right)$. Extremely small posterior predictive $p-$values indicate a clear rejection of the proposed model.

*Performance measures*

We perform the model selection using the DIC criterion (Spiegelhalter et al., 2002). Additionally, to compare the predictive performance we use (i) Average Absolute Prediction Error (AAPE) and (ii) Average Absolute Deviation (AAD) criteria. We use the posterior median as the point estimates of the predicted annual maxima for the validation set. The choice of posterior medians as point estimates seems to be natural because of the presence of skewness in the posterior predictive distribution. Let $\boldsymbol{M}_{i,t+1} = (M_{i,t+1,1},\ldots,M_{i,t+1,B})$ be a $B\times 1$ vector denoting the samples from the posterior predictive distribution for location $i$ at time $t+1$. Let $\tilde{M}_{i,t+1}$=Median$(\boldsymbol{M}_{i,t+1})$. Then, we define AAPE as

$$AAPE = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{T}|\tilde{M}_{i,t+j} - M_{i,t+j}|$$

where $M_{i,t+j}$ is the observed annual maximum corresponding to location $i$ in the year $t+j,\ \ j=1,2,\cdots$.

The AAD is further defined as

$$AAD = \frac{1}{nBT}\sum_{b=1}^{B}\sum_{i=1}^{n}\sum_{j=1}^{T}|M_{i,t+j,b} - \tilde{M}_{i,t+j}|.$$

## V.4. **Results**

We use the precipitation maxima of the year 1900 as the initial condition for estimation of $\boldsymbol{\beta}_0$. We train both the models on the dataset comprising of the annual maxima from the year 1901 through 1990 and validate it on the last eight years of data, i.e, from 1991 through 1998. As described in Section V.3, we first use the posterior predictive $p-$value to test the hypothesis about the goodness-of-fit of the models under consideration. If the $p-$ values do not indicate any lack of fit, we use DIC for model selection purpose and assess the predictive performance using the AAPE and AAD criteria. Table 9. shows DIC, AAPE and AAD scores along with the posterior predictive $p-$value for the Poisson-GP and the GEV models obtained for the proposed MEV copula and the baseline $t_5$ copula models. The posterior predictive $p-$value is fairly high for all the models under consideration indicating no lack of fit. However, for both the approaches, the $p$-values obtained for the MEV models are larger than that for the baseline model. Among the better performing MEV models, the DIC chooses the Poisson-GP model over the GEV model. Also note that, for both the approaches, the DIC prefers the MEV copula model to the baseline $t_5$ copula model. The predictive performance of the Poisson-GP model is better than the that of the GEV model for both the MEV and $t_5$ copula models. This indicates that it is more beneficial to consider the amount of exceedance in conjuction with its occurrence. In fact this finding agrees with the findings of Madsen et al. (1997) who conjectured that if other relatively high values in the sample were used besides the annual maxima, then more accurate estimates of the quantiles of the extreme value distributions would be obtained. However, note that, for both the approaches, the predictive performances of the MEV model are way better than that for the baseline model.

Table 9.: DIC, AAPE and AAD for the two competing models

| Copula | Models | $p_p$ | DIC | AAPE | AAD |
|---|---|---|---|---|---|
| MEV | Poisson-GP | 0.80 | 870.44 | 27.47 | 57.69 |
| | GEV | 0.71 | 874.14 | 38.93 | 64.64 |
| $t_5$ | Poisson-GP | 0.77 | 893.63 | 32.72 | 77.93 |
| | GEV | 0.71 | 900.97 | 41.27 | 82.24 |

The shape parameter $\boldsymbol{\alpha}$ is a key parameter that we need to draw inference on because the tail behavior of the marginal distribution depends exclusively on it. Table 10. shows the posterior summary of the shape($\boldsymbol{\alpha}$), scale($\boldsymbol{\rho}$) and dependence($\delta$) parameters for the Poisson-GP model and the GEV model. The posterior median of the shape parameter for both the regions under both models are negative. This indicates that the marginal distributions belong to the light tailed family. Note that, for both the models, the posterior median for $\boldsymbol{\alpha}$ corresponding to coastal regions is higher than that for the inland regions . As a matter of fact, the 95% posterior credible interval of the shape parameter in the coastal region for the GEV model contains 0, indicating a possibility that the marginal distribution of the annual maxima might as well belong to the very heavy-tailed Gumbel family. To get a clearer picture, we plot the kernel density estimates of the posterior distribution of $\boldsymbol{\alpha}$ for the inland and coastal stations for both the Poisson-GP and the GEV model in Figure 18. It is evident that, for both the models, the posterior distribution of $\boldsymbol{\alpha}$ for the coastal region is stochastically larger than its counterpart for the inland region. These facts clearly indicate the prevalence of relatively heavier tails at the coastal region. One of the meteorological reasons for this phenomenon is the seasonal development of tropical storms in the mid-Atlantic and Gulf of Mexico causing heavy precipitation at the eastern and south-eastern coast of USA. We have assumed the $\theta_1$ and $\theta_2$ to be *a pri-*

*ori* independent, however *a posteriori* they are dependent with correlation coefficient around 0.11 for the Poisson-GP model and 0.18 for the GEV model. Also note that the estimates of the dependence parameter ($\delta$) for the Poisson-GP model and the GEV model are similar with former yielding slightly lower estimates as compared to the latter.

Figure 19 shows the posterior median of $\boldsymbol{\beta}_t$ and 95% credible interval at each time point for the training set for the Poisson-GP model. Firstly, note that the coefficients corresponding to longitude are all positive and significant at most of the time points (Figure 19(b) ) indicating that for a given latitude-longitude coordinate of a site, the effect of the latter on extreme precipitation events is more significant as compared to the effect of the former. Secondly, note that negative association between elevation and extreme precipitation (Figure 19(c)).

Figure 20 shows the predictive maps of the point estimates (pointwise posterior predictive median) for the annual precipitation maxima produced by the relatively superior Poisson-GP model for the years 1991, 1995 and 1998. These maps show interesting geographic effects. The eastern and south-eastern regions show high intensity of extremal events as compared to the western and mid-western regions. The difference in the intensity of extreme precipitation between low-altitude regions and mountainous regions is also apparent. This finding is consistent with the negative association between elevation and extreme precipitation as described earlier and agrees with the findings of Jarrett (1990, 1993) who claimed that the hydrologic and paleohydrologic evidences show that intense rainfall does not occur at higher elevations. Also note the significant evolution of the random field over time. The predictive maps clearly show an increase in the occurrence of extreme precipitation events in 1998 as compared to 1991. Such increasing trends were also reported in Groisman et al. (2001), Kunkel et al. (1991) among others.

Table 10.: Posterior summary of parameters for the two competing models

| Models | Parameter | Median | 95% Credible Interval |
|---|---|---|---|
| Poisson-GP | Shape (coastal) | -0.40 | (-0.49, -0.21) |
| | Shape (inland) | -0.56 | (-0.80, -0.38) |
| | Scale | 0.46 | (0.02, 0.91) |
| | Dependence ($\delta$) | 7.96 | (1.18, 11.89 ) |
| GEV | Shape (coastal) | -0.02 | (-0.05, 0.09) |
| | Shape (inland) | -0.12 | (-0.27, -0.04) |
| | Scale | 0.41 | (0.05, 0.77) |
| | Dependence ($\delta$) | 12.55 | (4.17, 14.21 ) |

One of the main reasons to adopt Bayesian methodology was to obtain the uncertainty estimates associated with the predictions. Figure 21 shows the maps of uncertainty range, which is calculated by taking the difference between the pointwise 0.025 and 0.975 empirical quantiles from the posterior predictive draws, associated with predictive atlases shown in Figure 20. As one might expect, the uncertainty increases (range becomes wider) as the forecast horizon increases. Given a particular year, one can also see higher level of uncertainty at the desert locations and high altitude regions where few stations are location and the model is forced to extrapolate.

## V.5. **Sensitivity analysis**

There are three issues that are required to be addressed here:

(a) Choice of priors.

(b) Choice of the threshold.

*Choice of priors*

In a Bayesian analysis, the sensitivity of the results to the choice of the priors are required to ascertained. In the present study, the choice of priors for the dependence

parameter, $\delta$ and shape parameters $\boldsymbol{\alpha}$ are of paramount interest.

*Choice of dependence parameter*

A preliminary sensitivity analysis performed on various values of $\delta$ suggested that the model sensitive to the lower bound of $\delta$. When variables are strongly dependent, a large value of $\delta$ (typically $\delta \geq 10$, representing an upper tail-dependence in excess of 0.8 ) yields a posterior predictive $p$-value greater than 0.8 indicating a good model fit. But a low value of $\delta$ ($\delta \leq 1.2$, representing an upper-tail dependence lower than 0.5 ) yields a posterior predictive $p$-value of around 0.3 or less, indicating a somewhat questionable fit. The situation is reversed in case of near independent variables.

In practical situations, where the dependence between different pairs of the random variables may fluctuate widely, we suggest co-regionalization and choose different priors for $\delta$ for different regions.

*Choice of shape parameter*

The sign of the shape parameter determines the nature of the marginal distribution for both the models. Consequently, assumption of flat prior on the shape parameter makes the computation complicated and delays convergence of the chains. We have tested that if we can restrict the support of $\boldsymbol{\alpha}$ to either $\mathbb{R}^+$ or $\mathbb{R}^-$, the computation is far simpler and convergence is fast. However, in doing so, we need to be cautious about its impact on the prediction performance of the model. Simulation studies demonstrate that the prediction performance is worst under misspecification of the range of $\boldsymbol{\alpha}$. Hence we believe that it is better to sacrifice the computational simplicity in order to enhance the predictive accuracy of the model.

*Choice of threshold*

We perform a threshold sensitivity analysis by varying the return period, thereby changing the selected threshold, and then looking at how the shape parameter of the GPD changes with changes in threshold. It seems empirically that the posterior estimate of the shape parameter is most stable for low return periods (typically 1,2,3 years).

## V.6. **Concluding remarks**

We have developed a highly flexible Bayesian hierarchical model for analysis of spatio-temporal extremes. The model incorporates spatial dependence directly at the data level and admits unrestricted correlation structure. It has the ability to handle both strong and weak extremal dependences. To our knowledge, this is the first usage of such a flexible model to produce maps characterizing behavior of extremes across a geographic region and studying the evolution of the same over time. As an alternative to the usual block maxima approach we have empirically shown that it is more beneficial to consider exceedances over a certain threshold in order to come up with better predictive accuracy.

The statistical contribution of this work is twofold.

(a) A major drawback of the Bayesian approach to model extreme events is its poor performance when the variables are asymptotically dependent. This problem arises due to assumption of conditional independence of the data given the parameters. In addition to allowing incorporation of dependence directly at the data level, a major advantage of our model is that it can handle both asymptotically dependent and independent variables.

(b) Unlike most of the Bayesian models, which assume proper priors on dependence

parameters in order to guarantee posterior propriety, we have shown that under proposed model specifications, an improper prior on the dependence parameter, $\delta$, will lead to a proper posterior under mild regularity conditions.

The meteorological/hydrological contribution is the development of a methodology that can be employed to produce predictive and uncertainty maps of extreme precipitation over a geographical region. The methodology is superior to the commonly used regional frequency analysis algorithm in the sense that the latter cannot take into account all sources of uncertainty. In the posited model, on the other hand, the uncertainty arising from the parameter estimates as well as from the prediction procedure is accounted for.

The proposed model can be extended to a more generalized set-up. The assumption of fixed threshold can also be generalized by proposing location dependent thresholds. The assumption of stationarity in this case was directed by exploratory analyses. However, occurrences of annual and diurnal cycles are pretty common in hydro-meteorological events. The future challenge will be to incorporate such spatio-temporal non-stationarity in the the model.

# CHAPTER VI

# CONCLUSION

We have developed models for longitudinal, spatial and spatio-temporal processes in the Bayesian paradigm using various copulas to handle the dependences present in these processes. The introduction of copula in the data layer itself allows us to get rid of the conditional independence assumption. Although, the conditional independence assumption is commonly used to describe the data layer, it sometimes induces unwanted characteristics in the model - as we have shown in Chapter V. Essentially, copula method allows us to incorporate marginal dependence information directly into the model - a feature lacking in the conditional independence model. Such copula based hierarchical models have not been extensively studied in the Bayesian paradigm. So, the propriety of the posteriors are required to be investigated. In this dissertation, we have derived conditions required to obtain proper posteriors for specific choices of copulas. In addition, we have proved theoretically that the posited models do support the processes they are meant to model. Since we cannot observe a whole process in entirety, describing a paramteric model for the same always involves a chance of misspecification. We have performed extensive simulations to study the effect of such misspecifications on the posited models in Chapter IV. Besides studying these essential characteristics, we have obtained some remarkable features of the posited models. The multivariate bridge model presented in Chapter III has the ability of preserve interpretation of the fixed effects at the marginal scale after integrating out the random effects. The multivariate extreme value copula model presented in Chapter V can accommodate both strong and weak extremal dependence. Apart from these statistical

contributions, we believe that our models do have hydro-meteorological contributions as well. They can be employed to produce predictive and uncertainty maps of different hydro-meteorological phenomena over a geographical region. The methodology is superior to some commonly used models, like regional frequency analysis algorithm, in the sense that the former accounts for the uncertainty arising from the parameter estimates as well as from the prediction procedure. We believe that the development of these copula based Bayesian hierarchical models and the associated methodologies will go a long way to address the dearth of such models to study various geological and hydro-meteorological phenomena which yield both continuous and discrete outcomes.

# REFERENCES


Azzalini, A. and Capitanio, A. (1999) Statistical application of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, Series B*, **61**, 579–602.

Azzalini, A. and Dalla Valle A. (1996) The multivariate skew-normal distribution. *Biometrika*, **83**, 715–726.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. London: Chapman & Hall–CRC.

Barman, I., Sharma, A. K., Walker, R. F. and Datta-Gupta, A. (1998) Permeability prediction in carbonate reservoirs using optimal non-parmateric transformations: An application at the Salt Creek Field Unit, Kent County, TX, SPE 39667, presented at the 1998 SPE/DOE improved Oil Recovery Symposium, Tulsa, Oklahoma, April, 1998.

Cambanis, S., Huang, S. and Simons, G. (1981) On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**, 368–385.

Casson, E. and Coles, S. G. (1999) Spatial regression models for extremes. *Extremes*, **1**, 449–468.

Chen, M.-H., Ibrahim, J. G. and Shao, Q. M. (2004) Propriety of the posterior distribution and existence of the MLE for regression models with covariates missing at random. *Journal of the American Statistical Association*, **99**, 421–438.

Chen, M.-H. and Shao, Q. M. (2000) Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society*, **129**, 293–302.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004) *Copula Methods in Finance*. New York: Wiley.

Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Chilès, J.-P. and Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley.

Coles, S. G. (2001a) *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer-Verlag.

Coles, S. G. (2001b) Improving the analysis of extreme wind speeds with information-sharing models. *Notes de l'Institut Pierre Simon Laplace*, **11**, 23–34.

Coles, S. G. Heffernan, J. and Tawn, J. A. (1999) Dependence measuress for extreme value analyses. *Extremes*, **2**, 339–365.

Coles, S. G. and Tawn, J. A. (1991) Modelling multivariate extreme events. *Journal of the Royal Statistical Society, Series B*, **53**, 377–392.

Coles, S. G. and Tawn, J. A. (1996a) Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society, Series B*, **58**, 329–347.

Coles, S. G. and Tawn, J. A. (1996b) A Bayesian analysis of extreme rainfall data. *Applied Statistics*, **45**, 463–478.

Coles, S. G. and Tawn, J. A. (1994) Statistical methods for multivariate extremes: an application to structural design (with discussion). *Applied Statistics*, **43**, 1–48.

Coles, S. G. and Tawn, J. A. (2005) Bayesian modelling of extreme surges on the UK east coast. *Philos Transact A Math Phys Eng Sci*, **363**, 1387–1406.

Cooley, D., Nychka, D. and Naveau, P. (2007) Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, **102**, 824–840.

de Haan, L. and Resnick, S. I. (1993) Estimating the limit distribution of multivariate extremes. *Communications in Statistics–Stochastic Models*, **9**, 275–309.

Demerta, S. (2007) *The Copula Approach to Modelling Multivariate Extreme Values*. PhD Dissertation. Eidgenössische Technische Hochschule: Zurich.

De Oliveira, V. Kedem, B. and Short, D. S. (1997) Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association*, **92**, 1422–1433.

Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2002) *Analysis of Longitudinal Data*, 2nd edn. UK: Oxford University Press.

Dupuis, D. J. (2007) Using copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering* , **12**, 381–393.

Dupuis, D. J. and Tawn, J. A. (2001) Effects of misspecification in bivariate extreme value problems. *Extremes*, **4**, 315–330

Durman, C. F, Gregory, J. M., Hassell, D. C., Jones, R. G., and Murphy, J. M. (2001) A comparison of extreme European daily precipitation simulated by a global and regional climate model for present and future climates. *Queesland Journal of Royal Meteorological Society*, **127**, 573.

Easterling, D. R., Evans, J. L., Groisman, P. Y., Karl, T. R., Kunkel, K. E. and Ambenje, P. (2000) Observed variability and trends in extreme climate events: A brief review. *Bulletin of American Meteorological Society*, **81**, 417–425.

Eischeid, J. K., Diaz, H. F., Bradley, R. S. and Jones, P. D. (1991) A comprehensive precipitation data set for global land areas. *U.S. Department of Energy Report No. DOE/ER69017TH1*, Washington D.C.

Embrechts, P., McNeil, A. and Straumann, D. (2001) Correlation and dependency in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, (ed. M. Dempster, and H. Moffat), Cambridge University Press.

Fang, H.-B., Fang, K.-T. and Kotz, S. (2002) The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, **82**, 1–16.

Finkelstein, D.M., Williams, P.L., Molenberghs, G., Feinberg, J., Powderly, W., Kahn, J., Dolins, R. and Cotton, D. (1996) Patterns of opportunistic infections in patients with HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **12**, 38–45.

Fitzmaurice, G. M. (1995) A caveat concerning independence estimating equations with multivariate binary data. *Biometrics*, **51**, 309–317.

Fitzmaurice, G. M., Laird, N. M. and Rotnitzky, A. (1993) Regression models for discrete longitudinal responses. *Statistical Science*, **8**, 284–309.

Frees, E., and E. Valdez (1997) Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 1–25.

Genest, C., and Favre, A.-C. (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, **12**, 347–368.

Genton, M. G.(Ed.) (2004) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton: Chapman & Hall–CRC.

Geweke, J. (1989) Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **24**, 1317–1339.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C. and Razuvaev, V. N. (2001) Trends in intense precipitation in the climate records. *Journal of Climate*, **18**, 1326–1350.

Hall, P. and Tajvidi, N. (2000) Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, **6**, 835–844.

Hoel, P.G., Port, S.C. and Stone, C.J. (1971) *Introduction to Probability Theory*. Boston, MA: Houghton Mifflin Company.

Huerta, G. and Sansó, B. (2007) Time-varying models for extreme values. *Environmental Ecological Statistics*, **14**, 285–299.

Hulme, M., (1992) A 195180 global land precipitation climatology for the evaluation of GCMs. *Climate Dynamics*, **7**, 57–72.

Hulme, M., Osborn, T. J. and Johns, T. C. (1998) Precipitation sensitivity to global warming: Comparison of observations with HadCM2 simulations. *Geophysics Research Letters*, **25**, 3379–3382.

Jagger, T. and Elsner, J. (2006) Climatology models for extreme hurricane winds in the United States. *Journal of Climate*, **19**, 3220–3236.

Jarrett, R.D. (1990) Paleohydrologic techniques used to define the spatial occurrence of floods. *Geomorphology*, **3**, 181–195.

Jarrett, R.D. (1993) Flood elevation limits in Rocky Mountains. In *Engineering Hydrology: Proceedings of the Symposium Sponsored by the Hydraulics Division of the American Society of Civil Engineers*, (ed. C. Kuo), pp. 180–185. San Francisco, CA: American Society of Civil Engineers.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, H. and Hu, T. (1996) Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, **57**, 240–265.

Johnson, N. and Kotz, S. (1972) *Continuous Multivariate Distributions*. New York: Wiley.

Jorgensen, B. (1987) Small-dispersion asymptotics. *Brazilian Journal of Probability and Statistics*, **1**, 59–90.

Kim, H.-M. and Mallick, B.K, (2003) A note on Bayesian spatial prediction using the elliptical distribution. *Statistics and Probability Letters*, **64**, 271–276.

Kim, H.-M. and Mallick, B. K. (2004) A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning & Inference*, **120**, 85–101.

Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.

Kunkel, K. E., Andsager, K. and Easterling, D. R. (1999) Long-term trends in extreme precipitation events over the conterminous United States and Canada. *Journal of Climate*, **12**, 2515–2527.

Leadbetter, M. R., Lindren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer.

Leadford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika*, **83**, 169–187.

Leadford, A. W. and Tawn, J. A. (1997) Modelling dependence withing joint tail regions. *Journal of the Royal Statistical Society, Series B*, **59**, 475–499.

Lee, S. H. and Datta-Gupta, A. (1999) Electrofacies characterization and permeability predictions in carbonate reservoirs: Role of multivariate analysis and nonparametric regression, SPE 56658, presented at the 1999 SPE Annual Technical Conference and Exhibition, Houston, Texas, October, 1999.

Li, M., Boehnke. M., Abecasis, G. R. and Song,P. X.-K (2006) Quantitative Trait Linkage Analysis Using Gaussian Copulas. *Genetics*, **173**, 2317–2327

Madsen, H., Pearson, C. P. and Rosjberg, D. (1997) Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling. *Water Resources Research*, **33**, 759–770.

Moran, P. A. P. (1950) Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–33.

Natarajan, R. and McCulloch, C. E. (1995) A note on the existence of the posterior distribution for a class of mixed models for binomial responses . *Biometrika*, **82**, 639–643.

Natarajan, R. and McCulloch, C. E. (1998) Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics*, **7**, 267–277.

Natarajan, R. and Kass, R. E. (2000) Reference Bayesian methods for generalize linear mized models. *Journal of American Statistical Association*, **95**, 227–237.

Nelsen, R. B. (1999) *An Introduction to Copulas*. New York: Springer-Verlag.

NERC (1975) *The Flood Studies Report*. London: The Natural Environmental Research Council.

Nolan, J. P. (1997) Numerical calculation of stable densities and distribution functions. *Communications in Statistics–Stochastic Models*, **13** 759–774.

North, M. (1980) Time-dependent stochastic models of floods. *Journal of Hydraulics Division, ASCE*, **106**, 649–655.

O'Brien S., M. and Dunson, D. B. (2004) Bayesian multivariate logistic regression. *Biometrics*, **60**, 739–746.

Palacios, M. B. and Steel M. F. J. (2006) Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, **101**, 604–618.

Peddibhotla, S., Cubillos, H., Datta-Gupta, A. and Wu, C. H. (1996) Rapid simulation of multiphase flow through fine-scale geostatistical realizations using a new 3-D streamline model: A field example, SPE 36008, presented at Dallas, Texas, June, 1996.

Pickands, J. (1981) Multivariate extreme value distributions. In *Proceedings of the 43rd Sessions of the International Statistical Institute*, pp. 859–878. Buenos Aires.

Pitt, M., Chan, D. and Kohn, R. (2006) Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, **93**, 537–554.

Renard, B. and Lang, M. (2007) Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, **30**, 897–912.

Resnick, S. I. (1987) *Extreme Values, Point Processes and Regular Variation*. New York: Springer-Verlag.

Ryu, D., Sinha, D., Mallick, B., Lipsitz,S. R. and Lipshultz, S. E. (2007) Longitudinal studies with outcome-dependent follow-up: models and Bayesian regression. *Journal of American Statistical Association*, **102**, 952–961.

Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Processes*. New York: Chapman & Hall.

Sang, H. and Gelfand, A. E. (2009) Hierarchcial modeling for extreme values observed over space and time. *Environmental Ecological Statistics* (forthcoming).

Schon, J. H. (1996) *Physical Properties of Rocks: Fundamentals and Principles of Petrophysics*, New York: Pergamon Press.

Sklar, A. (1959) Fonctions de répartition à $n$ dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris*, **8**, 229–231.

Smith, R. L. and Naylor, J. (1987) A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution *Applied Statistics*, **36**, 358–369.

Smith, R. L., Tawn, J. A. and Coles, S. G. (1997) Markov chain models for threshold exceedances. *Biometrika*, **84**, 249–268.

Song, P. X.-K. (2000) Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, **27**, 305–320.

Spiegelhalter, D. Best, N., Carlin, B. and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

Stephens, M. (2000) Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B*, **62**, 795–809.

Stephenson, A. and Tawn, J. A. (2005) Bayesian inference for extremes: Accounting for the three extremal types. *Extremes*, **7**, 291–307.

Tawn, J. A. (1988) Bivariate extreme value theory: Models and estimation. *Biometrika*, **75**, 397–415.

Wang, Z. and Louis, T. A. (2003) Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. *Biometrika*, **90**, 765–775.

Zeger, S. L. and Liang, K.-Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.

Zeger, S. L., Liang, K.-Y. and Albert, P. S. (1988) Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, **44**, 1049–1060.

Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects; a Gibbs sampling approach. *Journal of American Statistical Association*, **86**, 79–86.

Zellner, A. (1976) Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, **71**, 400–405.

# APPENDIX A

# PROOFS OF RESULTS AND FIGURES

## PROOF OF RESULT 2.1

To show that the joint posterior is proper, we need to show $\int\int \pi(\beta, \Sigma|\mathbf{Y})d\beta d\Sigma < \infty$. Let us first show $\pi(\beta, \Sigma|\mathbf{Y}^*)$ is proper. Since $\mathbf{Y}^*$ consists of independent outcomes

$$\begin{aligned} \pi(\beta, \Sigma|\mathbf{Y}^*) &\propto \pi(\beta^|\mathbf{Y}^*)\pi(\Sigma) \\ &= L(Y^*|\beta)\pi(\Sigma) \ . \end{aligned}$$

Now

$$L(Y^*|\beta) = \prod_{i=1}^{n} \int (\text{pr}(\mathbf{Y}_i^* = y_i^*|\beta, b_i^*) f_b(b_i^*) \ ,$$

where $b_i^*$ is the random effect corresponding to the response $y_i^*$ and $f_b(.)$ is its univariate density function. Due to the marginal consistency property of the bridge random effects, we have

$$L(Y^*|\beta) = [F(\eta x_i'\beta)]^{y_i^*} [1 - F(\eta x_i'\beta)]^{1-y_i^*} \ . \tag{A.1}$$

Now, if $L(Y^*|\beta)$ has a unique maximum, then it is bounded. Then it follows from Theorems 2.1 and 3.1 of Chen and Shao (2000) that

$$\int L(Y^*|\beta)d\beta = M < \infty \ . \tag{A.2}$$

Now consider the entire data matrix $\mathbf{Y}$, so that the joint posterior is given by

$$\begin{aligned} &\int\int \pi(\beta, \Sigma|\mathbf{Y})d\beta d\Sigma \\ = \ &\int\int \prod_{i=1}^{n} \int \text{pr}(Y_{i1} = y_{i1}, \ldots, Y_{iT} = y_{iT}|\beta, \Sigma, \mathbf{b}_i) f_{\mathbf{b}}(b_{i1}, \ldots, b_{iT}) d\mathbf{b}_i \pi(\beta)\pi(\Sigma)d\beta d\Sigma \end{aligned}$$

$$
\begin{aligned}
&\leq \int\int \prod_{i=1}^{n} \int \mathrm{pr}(\mathbf{Y}_i^* = y_i^* | \beta, \Sigma, b_i^*) f_b(b_i^*) db_i \pi(\Sigma) d\beta d\Sigma \\
&= \int\int \prod_{i=1}^{n} \left[F(\eta x_i' \beta)\right]^{y_i^*} \left[1 - F(\eta x_i' \beta)\right]^{1-y_i^*} d\beta \pi(\Sigma) d\Sigma \\
&= \int L(\mathbf{Y}^* | \beta) d\beta \int (\pi(\Sigma) d\Sigma \quad [\mathrm{From}(A.1)] \\
&= M < \infty \quad [\mathrm{From}(A.2) \text{ and since the support of } \pi(\Sigma) \text{ is finite }]
\end{aligned}
$$

PROOF OF RESULT 3.1

For notational convenience we suppress the parameters of the marginal distribution of $Y_i$'s and denote the marginal distribution function and density function of $Y_i$ simply by $F_i(y_i)$ and $f_i(y_i)$ respectively.

The likelihood $L(\mathbf{Y}|\boldsymbol{\eta})$ can be decomposed into two parts, the so-called density weighing function

$$
\frac{g_{0,\Sigma}\left(Q_g^{-1}(F_{Y_1}(y_1;\mu_1,\rho_1,\alpha_1)), \ldots, Q_g^{-1}(F_{Y_n}(y_n;\mu_n,\rho_n,\alpha_n))\right)}{\prod_{i=1}^{n} q_g\left(Q_g^{-1}(F_i(y_i;\mu_i,\rho,\alpha))\right)}
$$

and the marginal densities of $\mathbf{Y}$. The first part is the copula density of an $n$ dimensional elliptical distribution and hence absolutely continuous. The marginal distribution of $\mathbf{Y}_i$ is absolutely continuous for each $i$ defined on the same support as that of of the copula density. Then $L(\mathbf{Y}|\boldsymbol{\eta})$ is essentially a convolution of absolutely continuous densities and hence is necessarily absolutely continuous.

By similar argument we can easily prove the absolute continuity of the $M$-component model.

PROOF OF RESULT 3.2

To check the Kolmogorov consistency conditions we prove the proposition in two parts.

*Symmetry under Permutation*

Let $p_1, p_2, \ldots p_n$ be any permutations of $1, 2, ..., n$. To show that $L(\mathbf{Y}|\boldsymbol{\eta})$ is permutation invariant, all we need to show is

$$\begin{aligned} & g_{0,\Sigma}\left(Q_g^{-1}(F_{Y_1}(y_1;\eta_1)),\ldots,Q_g^{-1}(F_{Y_n}(y_n;\eta_n))\right) \\ = \; & g_{0,\Sigma}\left(Q_g^{-1}(F_{Y_{p_1}}(y_{p_1};\eta_{p_1})),\ldots,Q_g^{-1}(F_{Y_{p_n}}(y_{p_n};\eta_{p_n}))\right) \end{aligned}$$

This is pretty obvious because elliptical kernels are symmetric in its arguments.
For the $M$ component model we can prove the consistency under permutation similarly because in this case, the model is a discrete mixture of elliptical copula with individual component being symmetric in its argument, the whole density function is symmetric in its argument.

*Dimensional Consistency*

*Case 1: Single Component Model*

Let $\mathbf{Y} = (y_1, \ldots, y_n) \sim G_n(\Sigma_n)$ with the density function given by

$$f(y_1,\ldots,y_n) = g_{0,\Sigma_n}\left(Q_g^{-1}\left(F_1(y_1)\right),\ldots,Q_g^{-1}\left(F_n(y_n)\right)\right)\prod_{i=1}^{n}\frac{f_i(y_i)}{q_g\left(Q_g^{-1}\left(F_i(y_i)\right)\right)} \quad \text{(A.3)}$$

To show that the dimensional consistency is preserved we need to show that the joint distribution of any $(n-1)$ dimensional vector follows $G_{n-1}(\Sigma_{n-1})$. Without loss of generality we derive the joint distribution of $\mathbf{Y}_{-n} = (Y_1, \ldots, Y_{n-1})$. Take the transformation $Z_i = Q_g^{-1}(F_i(y_i))$ in (A.3). Then we have $\mathbf{Z} = (Z_1, \ldots, Z_n) \sim EC_n(0, \Sigma_n, g)$. Define a $(n-1) \times n$ matrix $D = [I^{(n-1)\times(n-1)}, 0^{(n-1)\times 1}]$ of rank $n-1$. Then the joint distribution of $\mathbf{Z}_{-n} = (Z_1, \ldots, Z_{n-1})$ is same as the distribution of $D\mathbf{Z}$. By using property 4 described in section IV.1.2 we have $\mathbf{Z}_{-n} \sim EC_{n-1}(\boldsymbol{\mu}_{-n}, \Sigma_{-n})$ where $\boldsymbol{\mu}_{-n} = (\mu_1, \ldots, \mu_{n-1})$ and $\Sigma_{-n}$ is the matrix $\Sigma$ without the $n^{th}$ row and $n^{th}$

column. Now reverting the transformation from $Z_i$ to $Y_i$, we get the joint distribution of $\mathbf{Y}_{-n}$ as

$$f_{\mathbf{Y}_{-n}}(Y_1,\ldots,Y_{n-1}) = g_{0,\Sigma_{-n}}\left(Q_g^{-1}(F_1(y_1)),\ldots,Q_g^{-1}(F_{n-1}(y_{n-1}))\right)\prod_{i=1}^{n-1}\frac{f_i(y_i)}{q_g\left(Q_g^{-1}\left(F_i(y_i)\right)\right)} \tag{A.4}$$

Comparing the expression of $G_n(\Sigma_n)$ in (A.3) and (A.4) it is obvious $(Y_1,\ldots,Y_{n-1}) \sim G_{n-1}(\Sigma_{n-1})$. Note that, if we integrate out $Y_1,\ldots,Y_{n-1}$ we are left with the marginal distribution of $Y_1$ given by

$$\begin{aligned} f_{Y_1}(y_1) &= q_g\left(Q_g^{-1}(F_1(y_1))\right)\frac{f_1(y_1)}{q_g(Q_g^{-1}(F_1(y_1)))} \\ &= f_1(y_1) \end{aligned}$$

Clearly $Y_1 \sim G_1(\Sigma_1)$ whose density function is same as $f_{Y_1}(.)$

*Case 2: M-component Model*

We have the joint distribution of $y_1,\ldots,y_n$ given by

$$f_{\mathbf{Y}}(y_1,\ldots,y_n) = \sum_{j=1}^{M}\pi_j g_{0,\Sigma_j^{n\times n}}\left(Q_g^{-1}\left(F_1(y_1)\right),\ldots,Q_g^{-1}\left(F_n(y_n)\right)\right)\prod_{i=1}^{n}\frac{f_i(y_i)}{q_g\left(Q_g^{-1}\left(F_i(y_i)\right)\right)} \tag{A.5}$$

with $\sum_{j=1}^{M}\pi_j = 1$.

Take the transformation $Z_i = Q_g^{-1}(F_i(y_i))$. Then we have the joint distribution of $\mathbf{Z} = (Z_1,\ldots,Z_n)$ given by

$$f_{\mathbf{Z}}(z_1,\ldots,z_n) = \sum_{j=1}^{M}\pi_j g_{0,\Sigma_j^{n\times n}}(z_1,\ldots,z_n) \tag{A.6}$$

Then using the same argument used for the single component model, if we integrate out $Z_n$, then each of the $M$ components will be the density function of an $n-1$ dimensional elliptical distribution with mean 0 and covariance matrix $\Sigma_{j(-n)}$, $j =$

$1, 2, \ldots, M$, where $\Sigma_{j(-n)}$ is the matrix $\Sigma_j$ without the $n^{th}$ row and $n^{th}$ column. Thus the joint distribution of $(Z_1, \ldots, Z_{n-1})$ is given by

$$f(z_1, \ldots, z_{n-1}) = \sum_{j=1}^{M} \pi_j g_{0,\Sigma_{j(-n)}^{n-1\times n-1}}(z_1, \ldots, z_{n-1})$$

Now reverting the transformation we get the joint distribution of $(Y_1, \ldots, Y_{n-1})$ as

$$\begin{aligned} f_{\mathbf{Y}_{-n}}(y_1, \ldots, y_{n-1}) = & \sum_{j=1}^{M} \quad \pi_j g_{0,\Sigma_{j(-n)}^{n-1\times n-1}}(Q_g^{-1}(F_1(y_1)), \ldots, Q_g^{-1}(F_{n-1}(y_{n-1}))) \\ & \prod_{i=1}^{n-1} \quad \frac{f_i(y_i)}{q_g\left(Q_g^{-1}\left(F_i(y_i)\right)\right)} \end{aligned}$$

Also note that due to the generalized exponential parameterization of the covariance matrix, the marginal distribution of each $Z_i$ is $EC_1(0, 1, g)$. Hence the marginal density of each $Y_i$ is given by $f_{Y_i}(.)$ Thus the dimensional consistency is preserved.

PROOF OF RESULT 3.3

*Case 1: Single Component Model*

From (4.2) we observe that $\text{Corr}(Q_g^{-1}(F_i(y_i)), Q_g^{-1}(F_j(y_j))) = \rho_{ij} \;\; 0 \leq \rho_{ij} \leq 1$. Then using the Fang et. al (2002) result we get theat Kendall's correlation coefficient between $Y_i$ and $Y_j$ is given by

$$\tau_{ij} = 4E_{F_{ij}}(F_{Y_i,Y_j}(y_i, y_j)) - 1 = \frac{2}{\pi}\arcsin(\rho_{ij})$$

Since $0 \leq \rho_{ij} \leq 1$ and so is $\tau_{ij}$. We have $\tau_{ij}$ to be a monotone function of $\rho_{ij}$, hence isotropy is preserved.

*Case 2: M-component Model*

In this case, $\text{Corr}(Y_i, Y_j)$ does not have a closed form. However based on certain assumptions we get an approximate expression for the $\text{Corr}(Y_i, Y_j)$. Let $\text{E}(Y_k) = \mu_k$

and $\text{Var}(Y_k) = \sigma_k^2$. Now we define $Z_i = Q_g^{-1}(F_i(y_i))$ as we did earlier. Assuming $\sigma_k^2$ to be small we can invoke the small dispersion asymptotics (Jorgensen 1987) and following Song (2000) we can write

$$Y_k = F_k^{-1}(Q_g(Z_k)) = \mu_k + \sigma_k Z_k + o(\sigma_k) \tag{A.7}$$

From (A.6) we see that $\text{Corr}\left[Q_g^{-1}(F_k(Y_k)), Q_g^{-1}(F_l(Y_l))\right] = \text{Corr}(Z_k, Z_l) = \sum_{j=1}^{M} \pi_j \rho_{j(kl)} = \gamma_{(kl)}$ (say). Then using (A.7) and the above expression for $\gamma_{(kl)}$ we get

$$\sigma_{kl} = \text{cov}(Y_k, Y_l) \approx \sigma_k \sigma_l \gamma_{(kl)}$$

Thus the correlation between $Y_k$ and $Y_l$ is approximately proportional to a convex combination of $(\rho_{1(kl)}, \rho_{2(kl)}, \ldots, \rho_{M(kl)})$. Thus given the values of $(\pi_1, \pi_2, \ldots, \pi_M)$, $\sigma_{kl}$ only depends on the distance between the $k^{th}$ location and the $l^{th}$ location. Hence isotropy is preserved.

PROOF OF RESULT 3.5

Let $K_{ij},\ 1 \leq i < j \leq n$, be a bivariate copula. Let $H_1, \ldots, H_n$ be the univariate c.d.fs. Let $M$ be the distribution function of a positive random variable with Laplace transform $\psi$. Then following Joe and Hu (1996), we can write

$$\begin{aligned} &\int_0^\infty \prod_{1 \leq i < j \leq n} K_{ij}^{\alpha}(H_i, H_j) \prod_{i=1}^{n} H_i^{\nu_i \alpha} dM(\alpha) \\ &= \psi\left( - \sum_{1 \leq i < j \leq n} \log K_{ij}(H_i, H_j) - \sum_{i=1}^{n} \nu_i \log H_i \right) \end{aligned} \tag{A.8}$$

If we want to marginalize over $H_n$, i.e, if $H_n \to 1$, we get

$$\int_0^\infty \prod_{1 \leq i < j \leq n-1} K_{ij}^{\alpha}(H_i, H_j) \prod_{i=1}^{n-1} H_i^{(\nu_i + 1)\alpha} dM(\alpha)$$

$$= \psi\left(-\sum_{1\le i<j\le n-1}\log K_{ij}(H_i,H_j)-\sum_{i=1}^{n-1}(\nu_i+1)\log H_i\right) \tag{A.9}$$

Now (A.8) yields a valid copula with uniform marginals if $H_i(u_i)=\exp(-p_i\psi^{-1}(u_i))$ with $p_i=(\nu_i+n-1)^{-1}$. Taking this transformation, we get the joint distribution function of $U_1,\ldots,U_n$ from (A.8) given by

$$\begin{aligned}&\mathbb{C}(u_1,\ldots,u_n)\\ =\;&\psi\left(-\sum_{1\le i<j\le n}\log K_{ij}(\exp(-p_i\psi^{-1}(u_i)),\exp(-p_j\psi^{-1}(u_j)))+\sum_{i=1}^{n}\nu_i p_i\psi^{-1}(u_i)\right)\end{aligned} \tag{A.10}$$

After marginalizing over $U_n$, we get the joint distribution function of $U_1,\ldots,U_{n-1}$ from (A.9)as

$$\begin{aligned}&\mathbb{C}(u_1,\ldots,u_{n-1})\\ =\;&\psi\left(-\sum_{1\le i<j\le n-1}\log K_{ij}(\exp(-p_i\psi^{-1}(u_i)),\exp(-p_j\psi^{-1}(u_j)))+\sum_{i=1}^{n-1}(\nu_i+1)p_i\psi^{-1}(u_i)\right)\end{aligned} \tag{A.11}$$

Choosing $\nu_i=0,\quad \forall i$, i.e, $p_i=(n-1)^{-1}$, the bivariate copula $K_{ij}(u_i,u_j)=u_iu_j\exp\left(\left[(-\log u_i)^{-\Sigma_{ij}}+(-\log u_j)^{-\Sigma_{ij}}\right]^{-1/\Sigma_{ij}}\right),\Sigma_{ij}>0$ and the Laplace transform $\psi(s,\delta)=\exp(-s^{1/\delta}),\delta\ge 1$, (A.10) will lead to the family (4.24) and after marginalsation over $M_n$, (A.11) will lead to (4.25)

PROOF OF RESULT 3.6

Let

$$I \;=\; \int_1^\infty \frac{\tilde{\delta}^n}{\tilde{\delta}\prod_{i=1}^n m_i}F_{\boldsymbol{M}}(y_1,\ldots,y_n)d\delta,\;\;[\text{since }\pi(\delta)\propto 1]$$

$$
\begin{aligned}
&= \int_1^{\infty} \frac{\tilde{\delta}^n}{\tilde{\delta}\prod_{i=1}^n m_i} e^{-A(\delta)}, \ \text{[from 4.27]} \\
&= \frac{\tilde{\delta}^n}{\tilde{\delta}\prod_{i=1}^n m_i} \int_1^{\infty} e^{-A(\delta)} d\delta \ \text{[Using the regularity condition (i)]} \qquad \text{(A.12)}
\end{aligned}
$$

All we need to show is that, $I < \infty$.

Using Taylor series to expand $A(\delta)$ about its MLE $\delta_0$ and using condition (ii) we get

$$
\begin{aligned}
A(\delta) &\approx A(\delta_0) + (\delta - \delta_0)A'(\delta_0) + \frac{(\delta - \delta_0)^2}{2} A''(\delta_0) \\
&= A(\delta_0) + \frac{A''(\delta_0)}{2}\left[2A^*(\delta_0)\delta - 2A^*(\delta_0)\delta_0 + \delta^2 - 2\delta_0\delta + \delta_0^2\right] \qquad \text{(A.13)}
\end{aligned}
$$

where $A^*(\delta_0) = \frac{A'(\delta_0)}{A''(\delta_0)}$.

Then completing the square w.r.t $\delta$ in (A.13) we get

$$
\begin{aligned}
A(\delta) &\approx A(\delta_0) + \frac{A''(\delta_0)}{2}\left[\{\delta - (\delta_0 - A^*(\delta_0))\}^2 - A^*(\delta_0)^2\right] \\
&= A(\delta_0) - \frac{A''(\delta_0)}{2} A^*(\delta_0)^2 + \frac{A''(\delta_0)}{2}(\delta - \delta_0^*)^2 \qquad \text{(A.14)}
\end{aligned}
$$

where $\delta_0^* = \delta_0 - A^*(\delta_0)$.

Then combining (A.12) and (A.14) we get

$$
\begin{aligned}
I &\approx \frac{\tilde{\delta}^n}{\tilde{\delta}\prod_{i=1}^n m_i} \int_1^{\infty} e^{-\left[A(\delta_0) - \frac{A''(\delta_0)A^*(\delta_0)^2}{2} + \frac{(\delta - \delta_0^*)^2}{2[A''(\delta_0)]^{-1}}\right]} d\delta \\
&= \frac{\tilde{\delta}^n}{\tilde{\delta}\prod_{i=1}^n m_i} e^{-A(\delta_0) + \frac{A''(\delta_0)A^*(\delta_0)^2}{2}} \int_1^{\infty} e^{-\frac{(\delta - \delta_0^*)^2}{2[A''(\delta_0)]^{-1}}} d\delta \\
&= \frac{\tilde{\delta}^n}{\tilde{\delta}\prod_{i=1}^n m_i} \kappa_1 \times \kappa_2 \times \kappa_3
\end{aligned}
$$

Where

$$
\begin{aligned}
\kappa_1 &= e^{-A(\delta_0) + \frac{A''(\delta_0)A^*(\delta_0)^2}{2}} \\
\kappa_2 &= \left[\frac{2\pi}{A''(\delta_0)}\right]^{1/2} \\
\kappa_3 &= \left[1 - \Phi\left(\frac{1 - \delta_0^*}{A''(\delta_0)^{-1/2}}\right)\right]
\end{aligned}
$$

Note that $0 \leq \kappa_1 \leq 1$ and by condition (iii) $\max(\kappa_2, \kappa_3) < \infty$. So from (A.15), we get

$$\begin{aligned} I &\approx \frac{\tilde{\delta}^n}{\tilde{\delta}\prod_{i=1}^n m_i}\kappa_1 \times \kappa_2 \times \kappa_3 \\ &< \infty \end{aligned}$$

Then, given the hierarchical models described in Section IV.4.3 the propriety of the posterior distributions obtained in (4.33) are guaranteed by the propriety of the priors of the rest of the parameters.

PROOF OF RESULT 4.1

For simplicity, let us consider a bivariate set-up. Let $\boldsymbol{M}_i = (M_{i1}, \cdots, M_{iT})^T, \quad i = 1, 2$ denote these two time series of annual maxima observed at locations $S_1$ and $S_2$ respectively. Since we assume a conditional temporal independence at the very outset, we suppress the subscript $t$ and denote the distribution and density function of $\boldsymbol{M}_i$ by $F_i$ and $f_i$ respectively. Then a measure of extreme dependence is given by

$$\chi = \underset{u\to 1}{Lim}\, P(U_1 > u | U_2 > u) = \underset{u\to 1}{Lim}\, \frac{P(U_1 > u, U_2 > u)}{1-u}$$

where $U_i = F_i(m_i)$. Without loss of generality, we assume $M_i \sim GEV(0, 1, \alpha_i)$ with the density and distribution functions specified in (5.2) and (5.3). Then we get the quantile function as

$$F_i^{-1}(u) = \frac{(-\log(u))^{-\alpha_i} - 1}{\alpha_i} \tag{A.15}$$

$$f(M_1, M_2 | \alpha_1, \alpha_2) = \prod_{i=1}^{2} f(M_i | \alpha_i)$$

$$\chi = \underset{u\to 1}{Lim}\, \frac{P[M_1 > F_1^{-1}(u), M_2 > F_2^{-1}(u)]}{1-u}$$

$$
\begin{aligned}
&= \lim_{u\to 1} \int \frac{P[M_1 > F_1^{-1}(u), M_2 > F_2^{-1}(u)|\alpha_1,\alpha_2]}{1-u}\pi(\alpha_1,\alpha_2)d\alpha_1 d\alpha_2 \\
&= \lim_{u\to 1} \int \frac{P[M_1 > F_1^{-1}(u)]P[M_2 > F_2^{-1}(u)]}{1-u}\pi(\alpha_1,\alpha_2)d\alpha_1 d\alpha_2 \\
&= \lim_{u\to 1} \int \left[1-e^{-[1+\alpha_1 F_1^{-1}(u)]^{-1/\alpha_1}}\right]\left[1-e^{-[1+\alpha_2 F_2^{-1}(u)]^{-1/\alpha_2}}\right]\pi(\alpha_1,\alpha_2)d\alpha_1 d\alpha_2
\end{aligned}
\tag{A.16}
$$

Now using (A.15) we get

$$
1-e^{-[1+\alpha_1 F_1^{-1}(u)]^{-1/\alpha_1}} = 1-u \tag{A.17}
$$

Hence

$$
\chi = \lim_{u\to 1} (1-u)\int \pi(\alpha_1,\alpha_2)d\alpha_1 d\alpha_2
$$

Then under the condition that the joint distribution of $\alpha_1, \alpha_2$ is proper ($\int \pi(\alpha_1,\alpha_2)d\alpha_1 d\alpha_2 < \infty$) we have

$$
\chi = 0
$$

Hence a conditional independence assumption will lead to asymptotically independent class under a proper prior assumption.

FIGURES

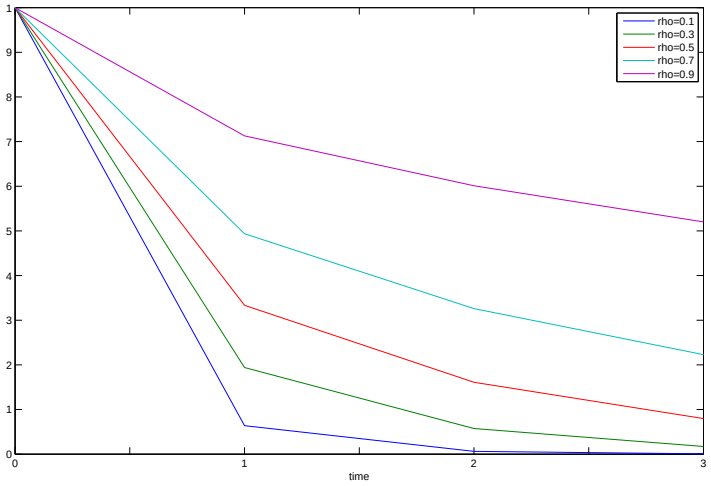
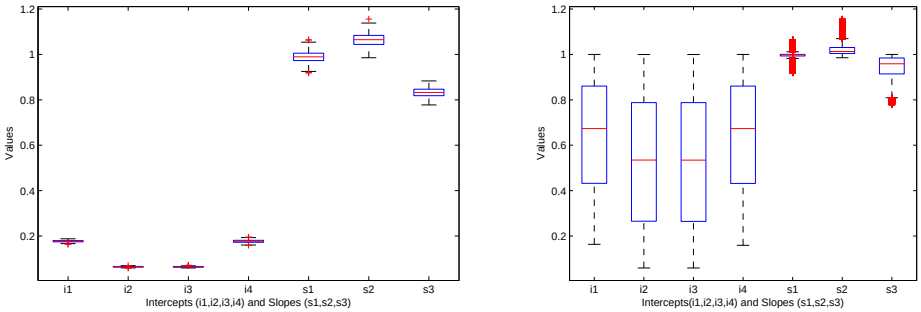
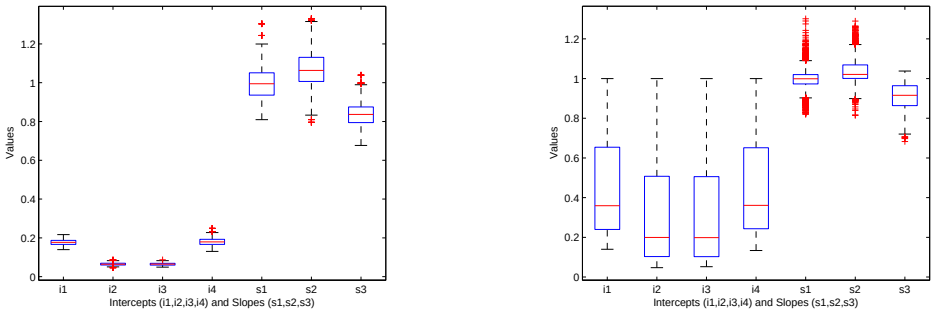



**Fig. 1**: Plot of Kendall's $\tau$ for two responses versus time-lag between observations for five values of $\rho$ and using $\rho_{st} = \rho^{|s-t|}$ for the parametric bridge random effects model.



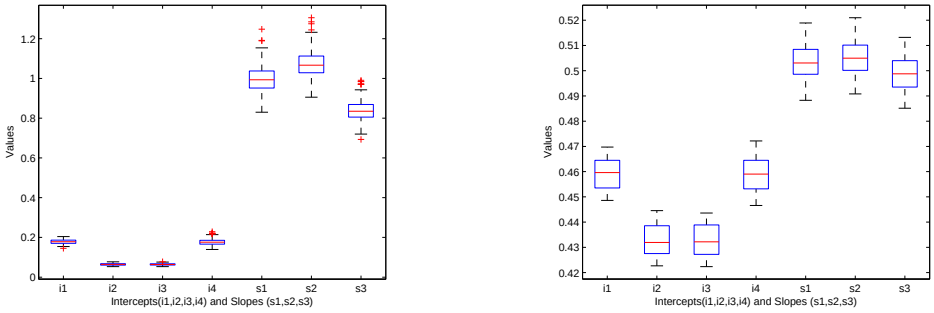

(a) Population-level odds (b) Marginal odds

**Fig. 2**: The posterior summaries of the regression parameters for logit-bridge Model



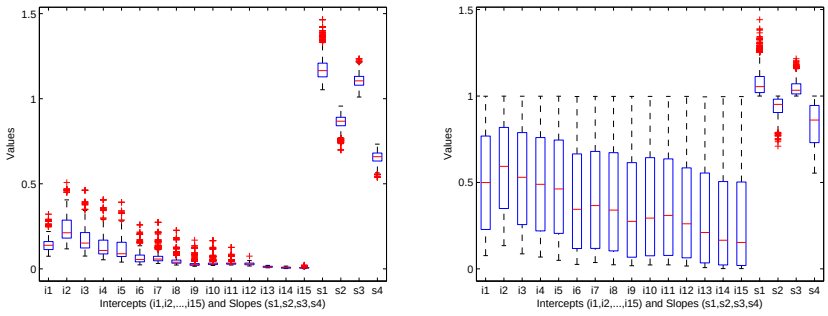

(a) Population-level odds (b) Marginal odds

**Fig. 3**: The posterior summaries of the regression parameters for log-log-stable Model




(a) Population-level odds (b) Marginal odds

**Fig. 4**: The posterior summaries of the regression parameters for logit-Gaussian Model




(a) Population-level odds (b) Marginal odds

**Fig. 5**: The posterior summaries of the regression parameters for semiparametric Model

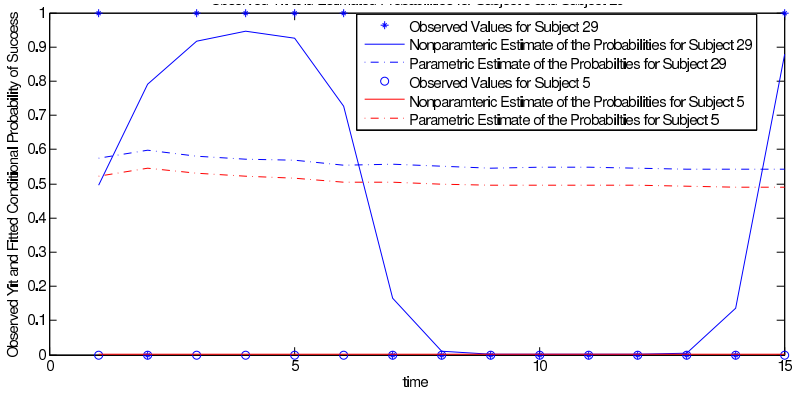



**Fig. 6**: Comparison between the predictive performance of the semiparametric bridge random effects model and the parametric bridge random effects model for Subject 5 and Subject 29.

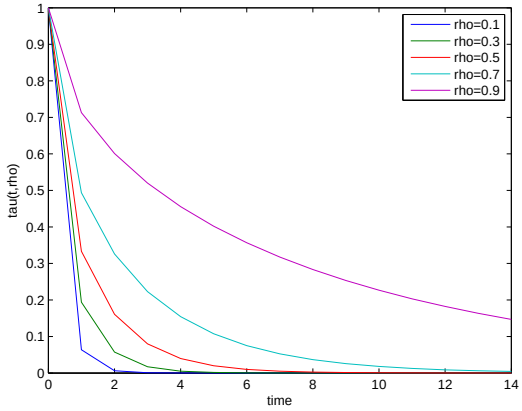
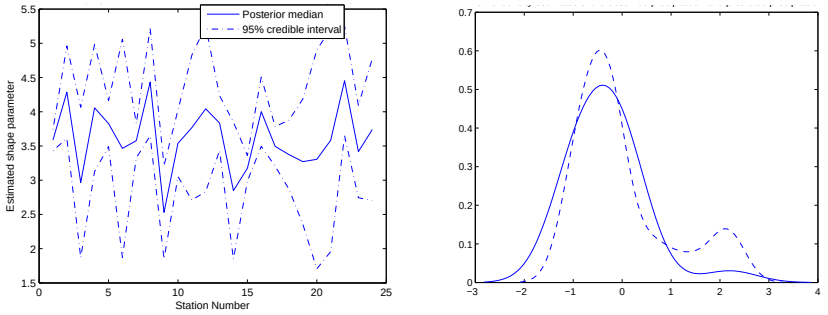



**Fig. 7**: Plot of Kendall's $\tau$ for two responses versus time-lag between observations for five values of $\rho$ and using $\rho_{st} = \rho^{|s-t|}$ for the semiparametric bridge random effects model.




(a) Posterior summary of $\alpha$ (b) Predicted distribution

**Fig. 8**: Posterior median (solid line) and 95% credible interval (dashed line) of shape parameter and Kernel smoothed predicted (solid line) and observed (dashed line) densities for Darwin data

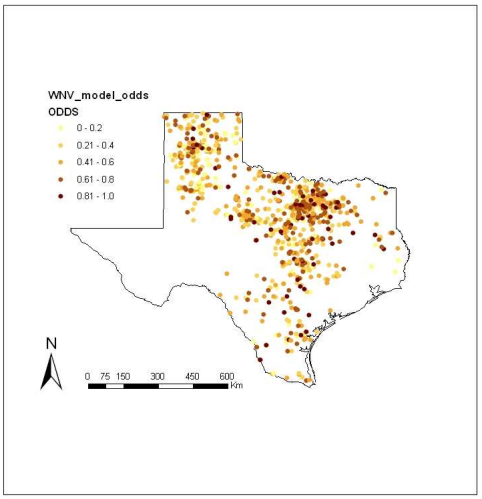
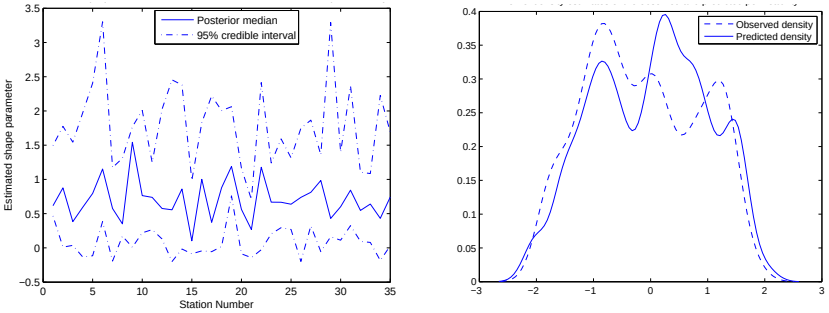
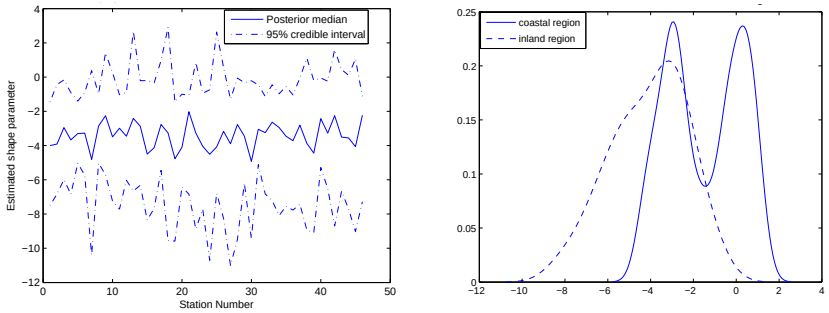



**Fig. 9**: Predicted odds of death from West Nile Virus at unobserved locations



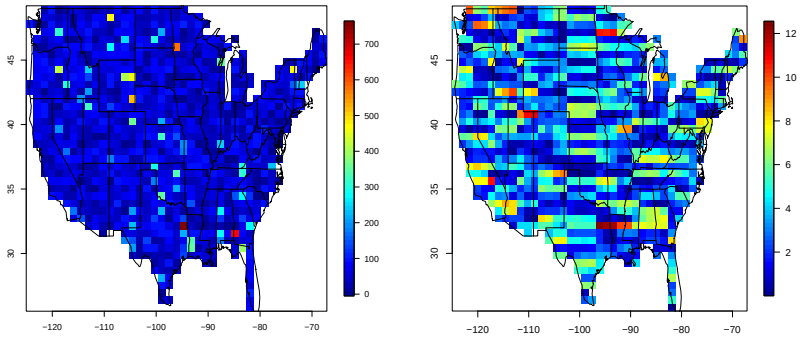

(a) Posterior summary of $\alpha$ (b) Pedicted distribution

**Fig. 10**: Posterior median (solid line) and 95% credible interval (dashed line) of shape parameter and Kernel smoothed predicted (solid line) and observed (dashed line) densities for permeability data




(a) Posterior summary of $\alpha$ (b) Posterior distribution of $\alpha$

**Fig. 11**: Posterior median (solid line) and 95% credible interval (dashed line) of shape parameter and Posterior distribution of the shape parameter for the coastal (solid line) and inland (dashed line) regions for the US maxima data



(a) Predictive map (b) Predictive uncertainty

**Fig. 12**: Extreme precipitation and Uncertainty map of US for the year 1998

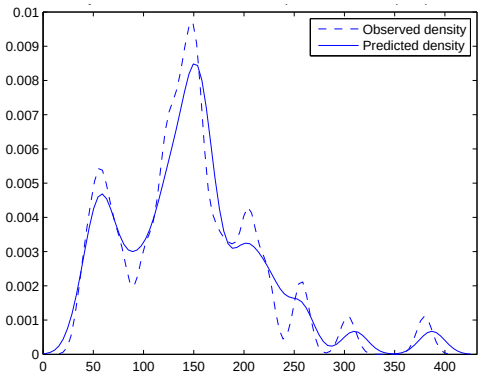



**Fig. 13**: Kernel smoothed predicted (solid line) and observed (dashed line) densities for U.S maxima data

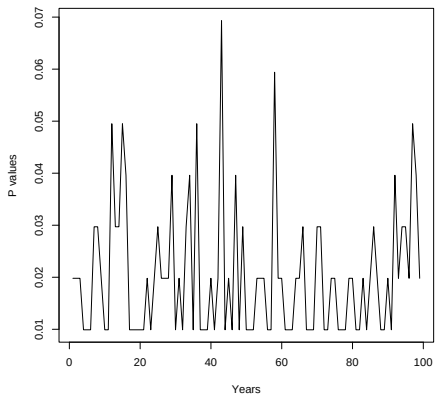



**Fig. 14**: Plot of the $p$ values for Moran's I randomization tests obtained at different time points

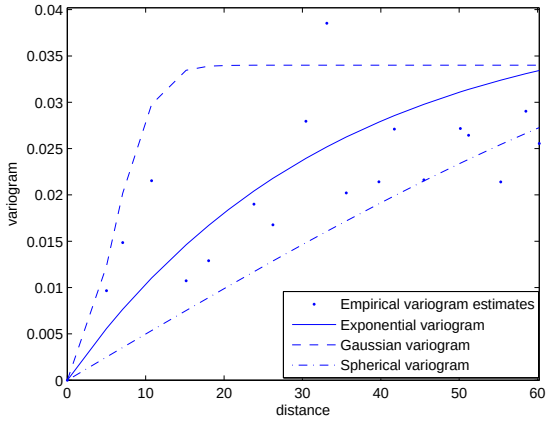
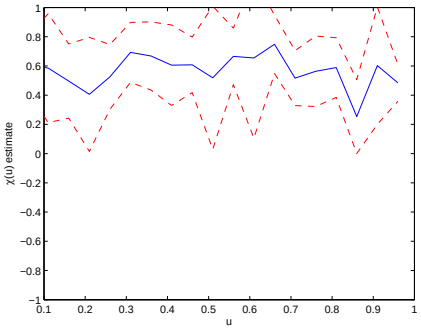
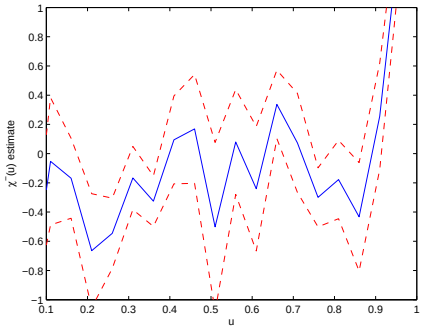


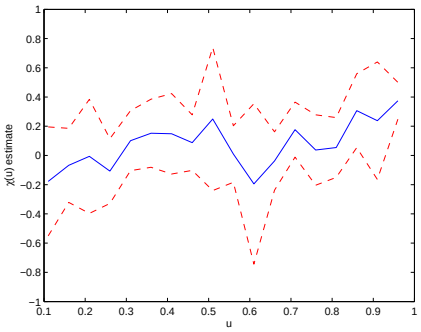

**Fig. 15**: Performance of the parametric variograms as compared to the Empirical binned variogram



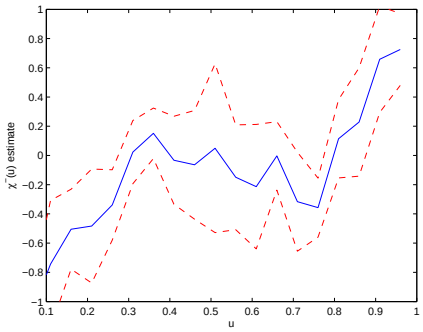

(a) $\chi(u)$ plot for pair 3 and 18




(b) $\bar{\chi}(u)$ plot for pair 3 and 18




(c) $\chi(u)$ plot for pair 10 and 13




(d) $\bar{\chi}(u)$ plot for pair 10 and 13

**Fig. 16**: Plots of estimated $\chi(u)$ and $\bar{\chi}(u)$ (solid line) and their approximate 95% confidence intervals (dashed line) for various stations

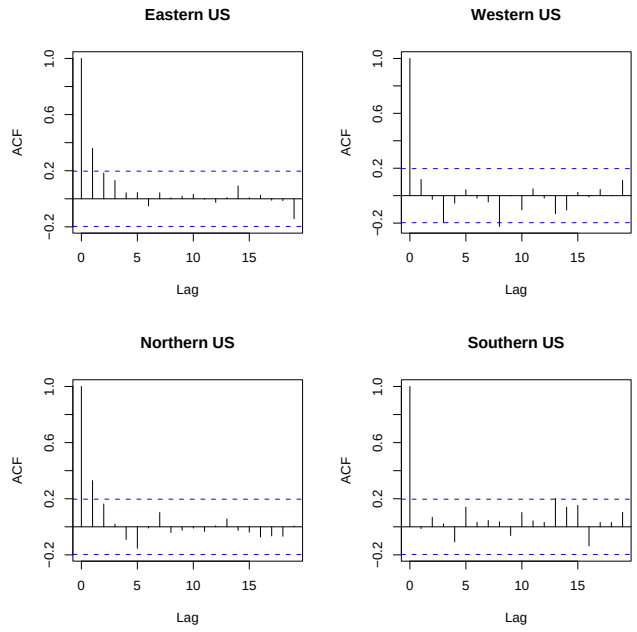



**Fig. 17**: ACF plot of the annual maxima time series observed at four different climatological regions

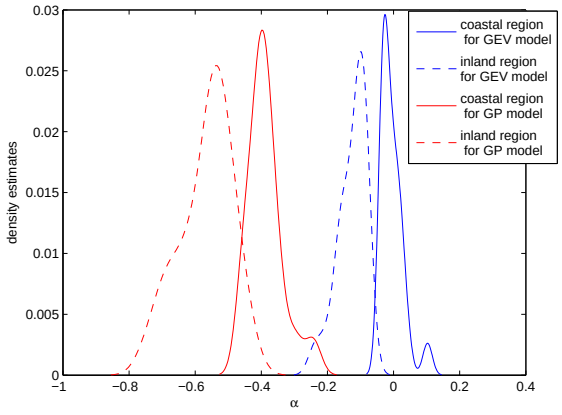
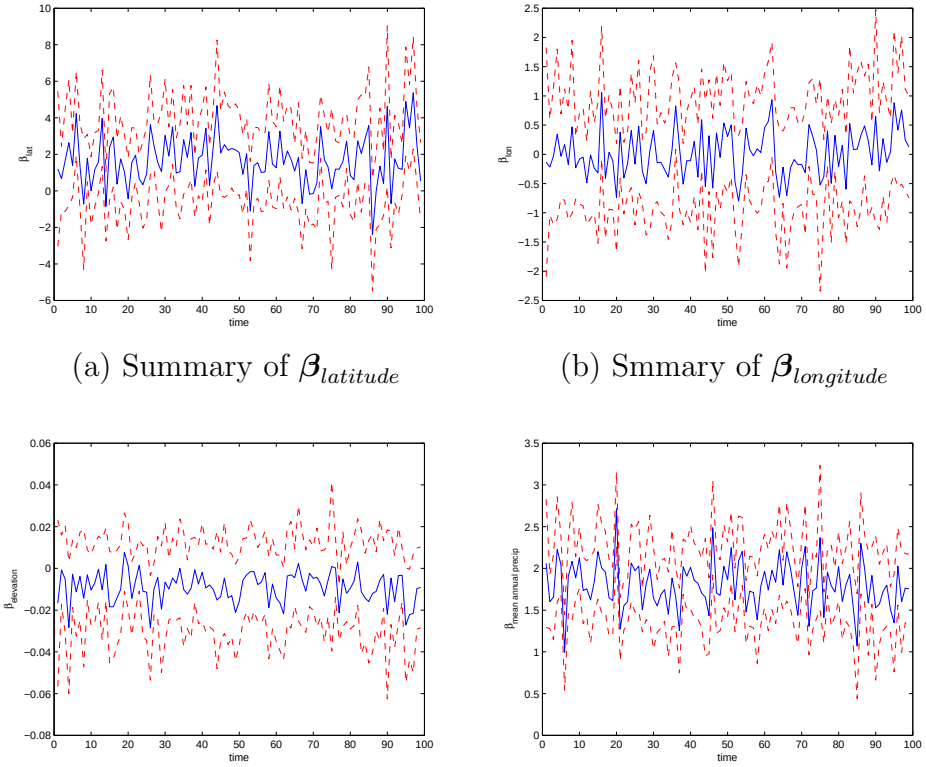


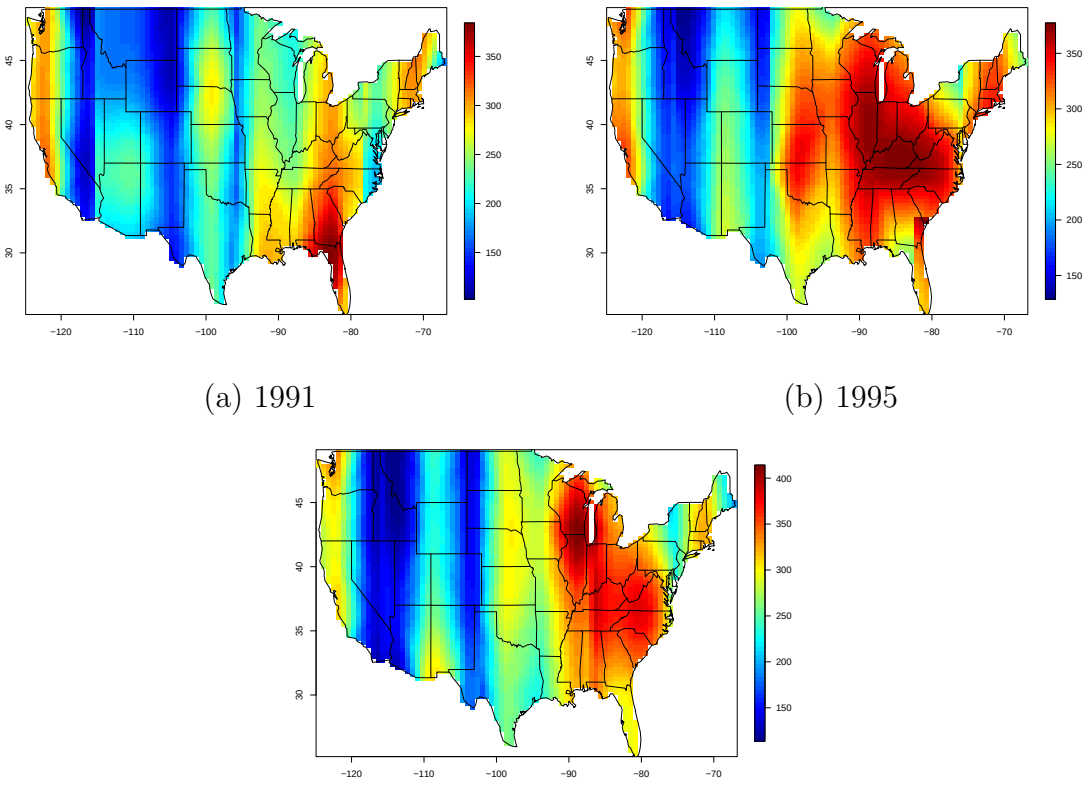

**Fig. 18**: Posterior distribution of the shape parameter for the inland and coastal regions under both models




(a) Summary of $\boldsymbol{\beta}_{latitude}$ (b) Smmary of $\boldsymbol{\beta}_{longitude}$

(c) Summary of $\boldsymbol{\beta}_{elevation}$ (d) Summary of $\boldsymbol{\beta}_{mean\ ann.\ precip.}$

**Fig. 19**: Posterior medians (solid line) and 95% credible intervals (dashed line) of the regression parameters

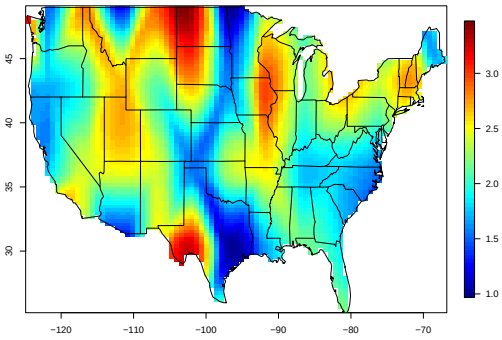



(a) 1991

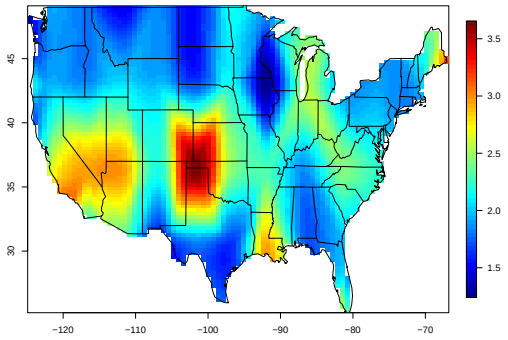
(b) 1995

(c) 1998

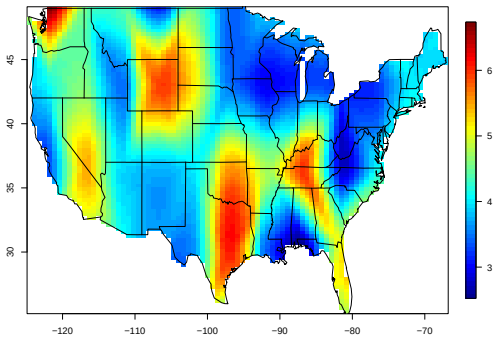
**Fig. 20**: Predictive maps of the point estimates for the annual precipitation maxima obtained using Poisson-GP model




(a) 1991




(b) 1995




(c) 1998

**Fig. 21**: Estimates of the range of 95% credible intervals for the predictive maps

# VITA

Souparno Ghosh

Department of Statistics

Texas A&M University

3143 TAMU

College Station, TX 77843-3143

c/o Bani K. Mallick, Ph.D.

## EDUCATION

2009 Ph.D., Statistics, Texas A&M University

2004 M.S., Statistics, University of Calcutta, India

2002 B.S., Statistics, University of Calcutta, India

## RESEARCH INTERESTS

Bayesian hierarchical models, Spatial and spatio-temporal modeling, Bayesian geostatistics.